DISCOVERY AND STRUCTURAL CHARACTERIZATION OF LANTIHPEPTIDES FROM SOIL AND RUMEN BACTERIA

BY

XILING ZHAO

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor Wilfred A. van der Donk, Chair
Professor William W. Metcalf
Associate Professor Douglas A. Mitchell
Professor Jonathan V. Sweedler

ABSTRACT

As members of the ribosomally synthesized and post-translationally modified peptide natural products (RiPPs), lanthipeptides possess myriad structural diversity and potential for discovery using genome-guided approaches. Biological activities of interest embedded within lanthipeptide structures include antimicrobial, antiviral, and morphogenetic functions. Continued discovery efforts seek to expand the characterized activities and structures of these compounds, as well as the enzymes that execute their biosynthesis.

The contents of this thesis are focused on the discovery of lanthipeptides using a variety of methods. Chapters 2 and 4 demonstrate the application of heterologous expression of gene clusters to produce lanthipeptides. The resulting peptide structures and enzyme activities provide intriguing insights into diversity of biosynthetic approaches present in this class of compounds. Furthermore, the abundance of precursor peptides encoded in the lanthipeptide biosynthetic systems discussed in Chapters 2 and 4 offer examples of combinatorial biosynthesis and the tantalizing possibility to harness this attribute. In contrast, Chapter 3 provides a glimpse into the targeted screening of *Actinobacteria* for the production of lanthipeptides, which suggests additional structural diversity. The successful application of these methods increases the number of lanthipeptide compounds and systems, which ultimately allows for a deeper understanding of and appreciation for this class of natural products.

For my family and friends

ACKNOWLEDGEMENTS

First and perhaps most deserving of credit for my completion of the PhD program at UIUC is my advisor Wilfred. I used to call you Professor van der Donk but you graciously allowed me to use your first name and pursue my PhD studies in your laboratory. Your consistent display of patience and kindness were invaluable in all of my graduate school endeavors. In addition, thank you for imparting your approach to science; working on science in your lab has been eye-opening and exceptionally educational.

All of the van der Donk lab members, past and present, have thoroughly instructed me in the application and utility of wisdom and patience. Meeting and working with the talented individuals in the van der Donk lab has been humbling. I am especially indebted to several lab member whom I highlight in no particular order. The Magnificent Seven, after you guys left I could only aspire to provide the supportive environment that you guys supplied. I hope I haven't embarrassed you guys too much with my outlandish stories of your "legendary" exploits. Manny, I am grateful for your offer to teach me how to run PCRs and basically every lab technique that I know and then for putting up with me. Chantal, you were the comical side to Manny's sometimes too serious disposition. I look forward to working with you in the future. I am also grateful for Dr. Walker and your wise Chihuahua eyes and words, and Kenton for your Varys-like approach to interpersonal relationships. Chang, Dr. Repka, Linna, and Dr. Yu the people who were in closest physical proximity to my lab desk, bench space, and temper tantrums. Thank you for being so gracious and understanding. Nick, thanks for tolerating my teasing and returning it in equal measure. Weixin, it was always a delight to see your audaciously colored pants and listen to your equally bold statements on science and life. Dr. Splain, I appreciate your holistic and pragmatic wisdom that started the first day I met you, which was also the first day I joined the lab. Dr. Funk, I admire your patience and determination to not only sign up for a really difficult project but to also stick it out. Dr. Huo, thanks for providing one of the most efficient and enjoyable scientific collaboration experiences I've had, due in no small measure to your experience and forbearance. To everyone, it's been a pleasure working and interacting with you, thank you for putting up with me.

Thank you to my committee members who were patient and instructive during each of my graduate milestones. I am privileged to have worked with several collaborators outside of the van der Donk lab. I thank to Professor Metcalf and Dr. Doroghazi for showing me the true

meaning of genome mining and enduring my rediscoveries. Professor Mackie and Dr. Kwon, thank you both for showing and teaching me the intriguing facets of anaerobic culture technique.

Thank you to all members of the UIUC staff who provided instrument and emotional support. Again, I want to specifically mention some of these people. It was a pleasure interacting with the secretaries in the chemistry department. You are all beyond efficient and friendly, scheduling my presentations would have been impossible without you guys. I am grateful for every member of the mail room for being prompt and helpful with shipping-related and unrelated questions I had. Hodge, thank you for helping me saw things in half and listening to my inane questions about 3-D printers. Dr. Ulanov, thanks for answering my rudimentary questions and for trusting me enough to run the GC/MS with minimal supervision. Huge thanks to Dr. Zhu for not only helping me with every phosphorus NMR experiment but also with signing me up for instrument time. Although I didn't have the pleasure to train with you, thank you Dr. Olson for always having something kind and charming to say, these were bright moments in sometimes dark days.

Before graduate school at UIUC, there was the Dorrestein lab. I have fond memories of my time in the lab and can say that some of the glimpses into graduate life that all of you gave me were spot on. Thank you all for calibrating my expectations for graduate school and supporting my decisions. Before all of this, there was and is my family, I wouldn't be here without your support. I am grateful for both my parents and their decision to grant me freedom in many senses of the word. Also, many thanks to my younger sister who I think is wiser than me in many ways and always listened to my concerns. Finally, I feel incredibly privileged to have met my best friend and significant other, Brian, at UIUC. Thank you for not running for the hills and putting up with my histrionic stories. The past is in the past and the future is with you.

TABLE OF CONTENTS

Chapter 1. Introduction

1.1 Ribosomally synthesized and post-translationally modified peptide natural products (RiPPs)

Natural products are a significant source of pharmaceutical compounds as well as tools to probe biological processes (1, 2). Recently, peptide natural products have attracted attention for the advantages they offer including high target selectivity and the potential to interact with targets with larger surface areas such as in protein-protein interactions. Despite these attractive features, several challenges remain including low metabolic stability (5). Natural products containing amide bonds can be stabilized by additional chemical groups that decorate the peptide scaffold (Figure 1.1). One set of compounds that exemplify this approach are ribosomally synthesized and post-translationally modified peptide natural products (RiPPs) which are a sprawling and growing class of compounds that can be subdivided based on their post-translational modifications (6). In contrast to non-ribosomal peptide synthetases, the scaffold for the final compound of a RiPP is encoded within the biosynthetic gene cluster along with the modification machinery. Within the RiPP designation are lanthipeptides, which exhibit an array of structural features (Figure 1.1, 1.2) as well as biological activities, with the subset of lanthipeptides that exhibit antimicrobial activity termed lantibiotics. Their post-translational modifications are key to the activity of lanthipeptides, thus there is sustained interest in exploring their biosynthesis as well as discovering new lanthipeptides.

Similar to other RiPPs, lanthipeptide biosynthesis is comprised of three parts: expression of the precursor peptide and protein genes, post-translational modification installation, and leader peptide removal (Figure 1.2). First, both the synthetase(s) and substrate (generically termed LanA) genes are transcribed and translated. The latter comprises an N-terminal leader peptide and a C-terminal core peptide, which are delineated by a sometimes conserved cleavage motif. The synthase enzyme(s) then catalyze(s) the dehydration and cyclization of core peptide residues in the substrate to dehydro amino acids (abbreviated as Dha and Dhb) and methyl(lanthionine) residues, abbreviated as (Me)Lan (Figure 1.2, 1.3). Post-translational modifications aside from the thioether rings can also be installed, and finally proteolytic removal of the leader peptide yields the mature lanthipeptide. The ribosomal heritage of lanthipeptides offers an opportunity

for facile modifications of the final product scaffold via mutagenesis, and additional modifications increase the chemical diversity of lanthipeptides.

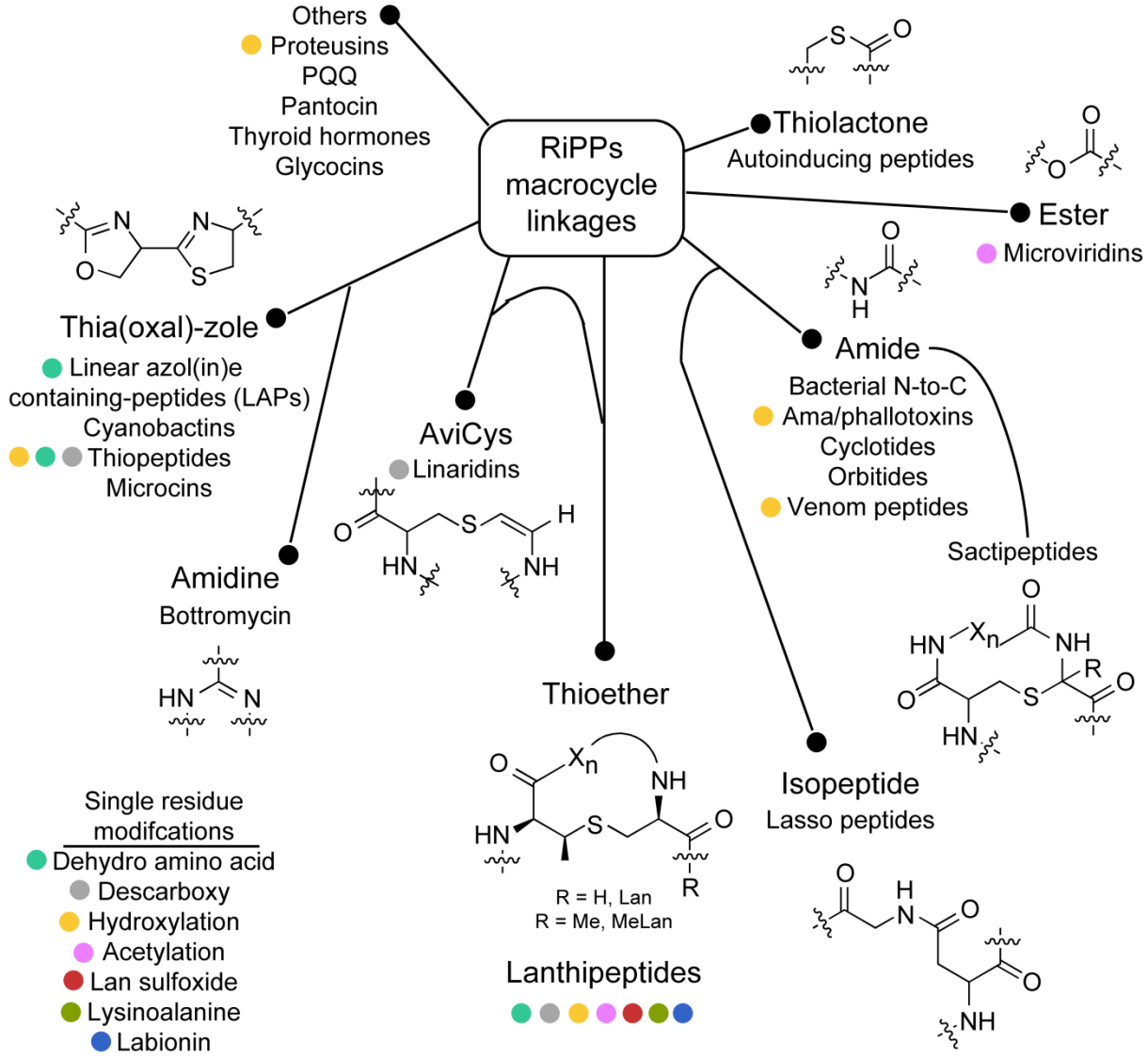

Figure 1.1. Chemical linkages found in RiPP macrocycles. Representative substructures are drawn. Additional modifications found in lanthipeptides, and shared with other classes, are indicated with colored circles. Shared structures are not installed by the same enzymes. Some macrocycle linkages are not exclusive for representative compounds, for example, sactipeptides also contain thioether linkages.

In addition to the namesake (Me)Lan residues that are key to their bioactivity, lanthipeptides possess additional post-translational modifications (Figure 1.1). Unlike the (Me)Lan linkages, these modifications usually occur at a single residue in the peptide. Investigation into the activities of the enzymes responsible for these modifications suggests substrate tolerance at an amino acid level, and in some cases enzyme activity in the absence of leader peptide. Thus, there is interest in developing these enzymes as substrate-tolerant bioengineering tools. Continued efforts toward the discovery of new lanthipeptides have revealed new post-translational modifications as well as new activities of the resulting peptide products (7-9).

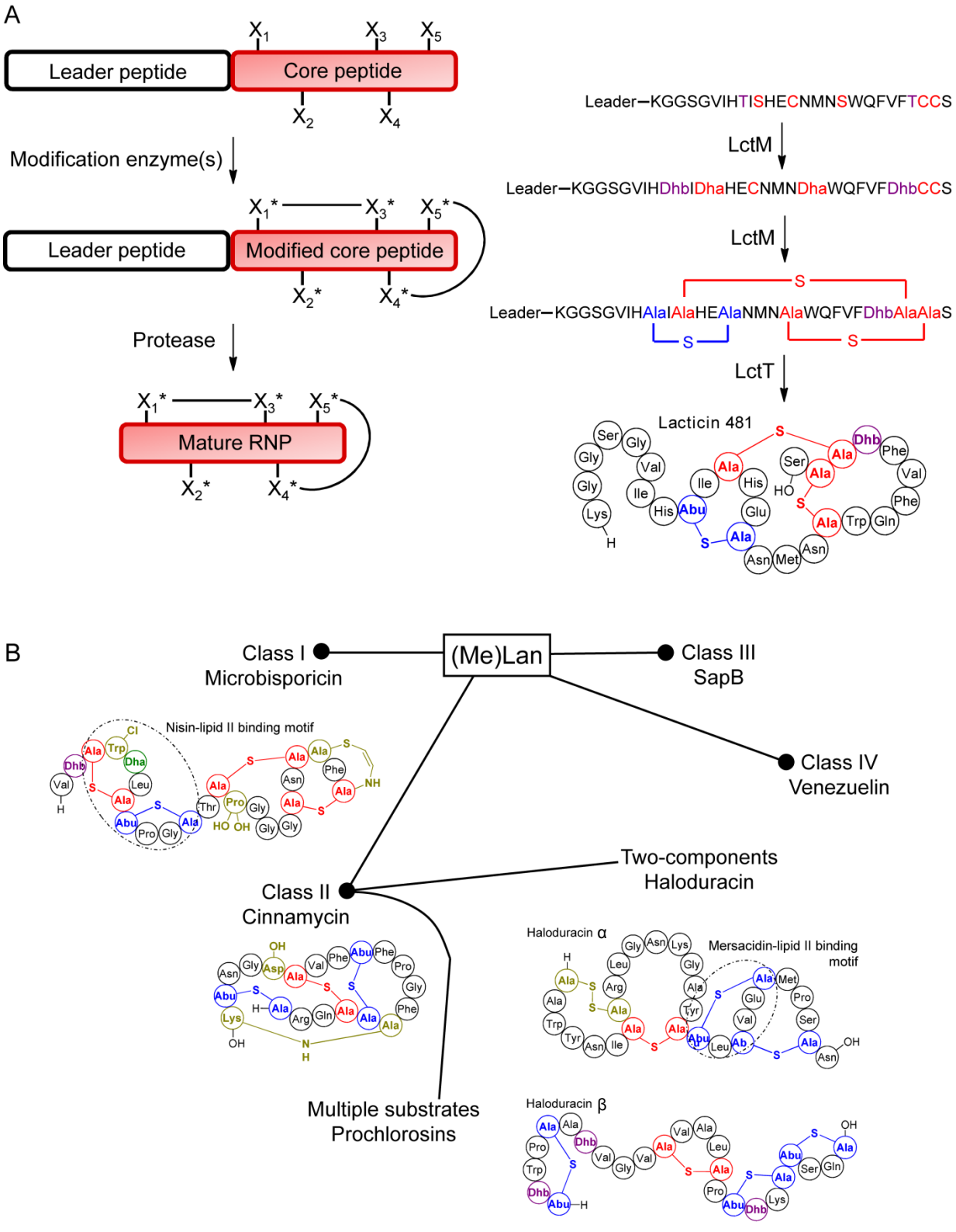

Figure 1.2. Lanthipeptide biosynthesis. (A) Lacticin 481 maturation shown in context of an outline of RiPP biosynthesis. (B) Structures of a variety of lanthipeptides. Lan shown in red, (Me)Lan in blue, and single residue post-translational modifications in chartreuse.

1.2 LanM proteins: structural and mechanistic insights

Lanthipeptides are divided into classes based on the bioinformatic annotation and mechanistic distinction of their synthases (Figure 1.3A). Compounds from class I are biosynthesized by a discrete LanB and LanC protein, class II are modified by a bifunctional LanM protein, class III by a trifunctional LanKC protein, and class IV by yet another trifunctional LanL enzyme (10). The LanB proteins in class I synthases are responsible for dehydration of the Ser and Thr residues within the LanA core peptide while the LanC cyclases catalyze the addition of the thiols of Cys residues onto the dehydro amino acids (Figure 1.3B, C). In comparison, the class II-IV synthetases rely on a single enzyme for both dehydration and cyclization activity. However, the class II-IV enzymes have distinct domains. The class II enzymes comprise an N-terminal dehydratase and C-terminal cyclase domain while class III and IV enzymes consist of a kinase, lyase and cyclase domain (Figure 1.3A). As a reflection of their annotated differences, the four classes of lanthipeptide synthases employ distinct strategies to affect the dehydration of core peptide residues (Figure 1.3B, C). The class I synthase NisB activates the hydroxyl of Ser and Thr residues by glutamylation, using aminoacylated glutamyl-tRNA as the glutamyl source (11, 12). In contrast, class II-IV enzymes use nucleoside triphosphates to phosphorylate Ser and Thr residues, which upon phosphate elimination produce dehydro amino acids (10).

Canonical LanM proteins are typically about 1,000 amino acids in length and can be divided into an N-terminal dehydratase domain and C-terminal cyclization domain (Figure 1.3A, 1.4). These domain borders can be visualized *in silico* using Pfam's Hidden Markov model (HMM)-based search, and the National Center for Biotechnology Information's (NCBI's) BLAST also outputs a cyclization domain annotation (13, 14). Alignment of characterized LanM proteins, which typically share around 30% identity, reveals the presence of conserved residues in both domains (10). Initial studies into the significance of these residues were carried out on the synthase LctM and necessitated reconstitution of its in vitro activity, which required both adenosine triphosphate (ATP) and Mg^{2+} (15). Subsequent mutations of conserved residues in the dehydratase domain of LctM resulted in mutant enzymes with decreased dehydration activity or that produced phosphorylated LctA. These observations suggested a mechanism in which net dehydration was achieved via addition of a phosphoryl group, from ATP, followed by phosphate elimination, with Mg^{2+} proposed to assist in elimination (Figure 1.3B) (16). Complementary

mutagenesis studies were undertaken for proposed Zn^{2+} coordinating residues (His725, Cys781, and Cys836) within the cyclization domain of LctM. Specifically, mutagenesis resulted in enzymes that were unable to cyclize the substrate, suggesting a thiol coordinating and activating role for zinc (Figure 1.4) (17).

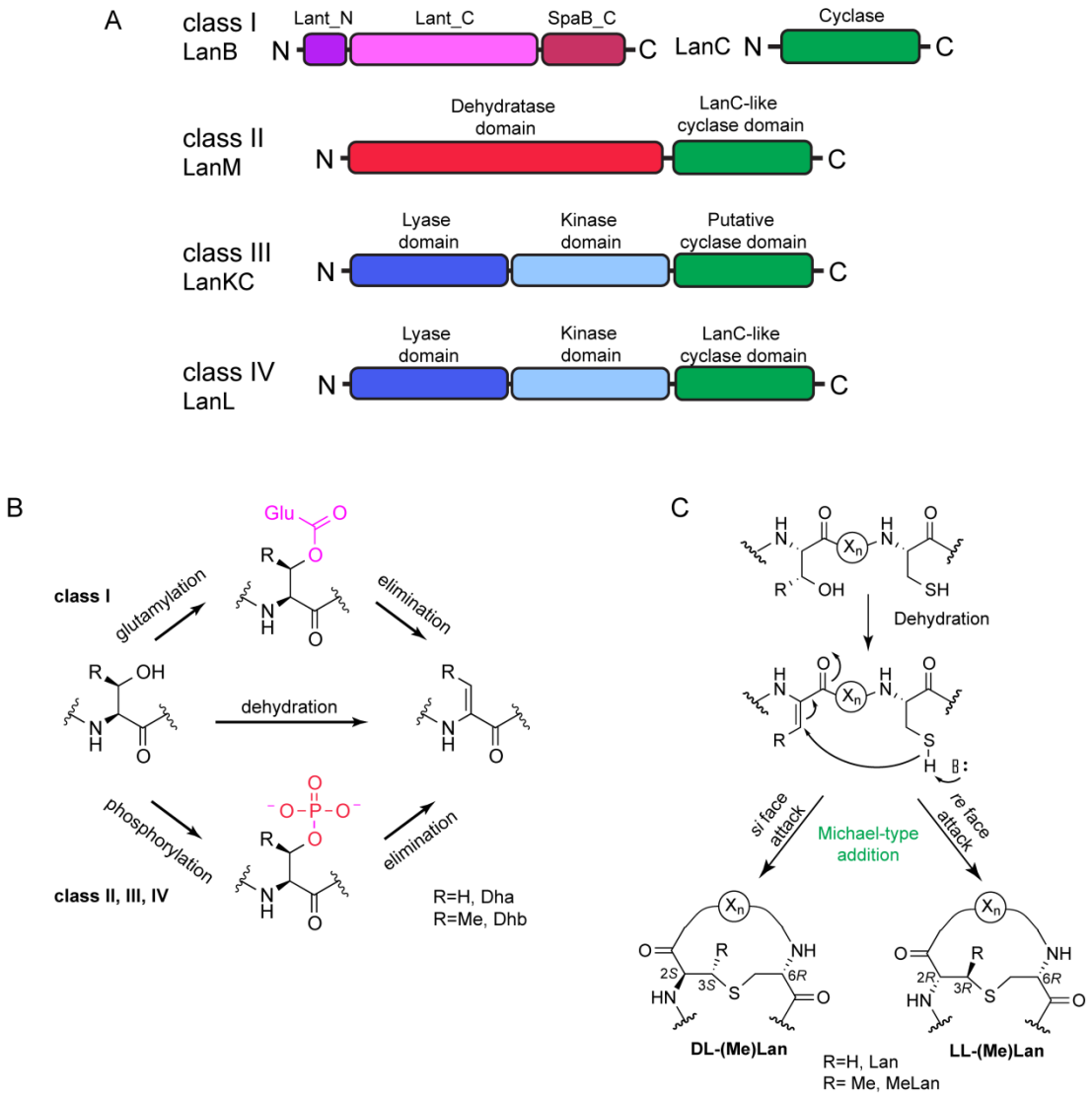

Figure 1.3. Lanthipeptide synthases and modifications. (A) Schematic of the domains of lanthipeptide synthases from each class. (B) Mechanism for dehydro amino acid formation. (C) Mechanism for thioether ring formation and resultant (Me)Lan stereochemistries. Figure adapted from (3).

These functional and structural predictions provided by the primary sequence and biochemical studies can be contextualized in the crystal structure of CylM. The CylM monomer consists of a dehydratase domain that resembles lipid kinases and which is attached to a C-terminal cyclization domain which contains a coordinated zinc. In addition to substantiating the proximity of residues identified as conserved by alignment, co-crystal structures with adenosine monophosphate (AMP) identified residues within a nucleotide binding pocket, with Asp252, His254, Arg506, and Thr512 proposed as involved in phosphate elimination (18). Inactivation of elimination activity via mutations at these residues was demonstrated for LctM, ProcM, and CylM, with the former two demonstrating potential application as general kinases as a result of their substrate tolerance (Figure 1.4) (19, 20). Due to lack of co-crystallization with substrate, the contacts that the leader peptide makes with the enzyme and the trajectory of the core peptide during catalysis remain elusive. The bioinformatic and structural separation of the dehydratase and cyclase domains could be translated to functional decoupling of the dehydratase and cyclase domain activities (Figure 1.4). Separation of the dehydration and cyclization domains of several LanM enzymes including BovM, CylM, LctM, NukM, and ProcM yielded outcomes ranging from fully functional to completely inactive discrete domains (21-25).

Studies employing time dependent assays, tandem mass spectrometry (MS), and selective Cys alkylation agents have provided some insight into the dehydration and cyclization reactions carried out by LanM enzymes (26). Initial studies employing high-resolution tandem MS and mutant LctM enzyme and HalM2 revealed distributive and an overall N- to C-terminal directional behavior for both enzymes (27). Subsequent semi-quantitative MS strategies have yielded higher resolution kinetic insight into the dehydration and cyclization reactions of wild type HalM2 and ProcM (28). The data substantiate the distributive behavior of these enzymes and provided evidence for alternating installation of dehydrated amino acids and thioether linkages with a general directionality, as opposed to full dehydration of the substrate followed by cyclization (28). Despite belonging to the same class of lanthipeptide synthetases, the overall directionality for ProcM is C- to N-terminal, which is opposite of that observed for HalM2. Furthermore, comparison of HalM2 and ProcM revealed substantial kinetic differences, with the latter exhibiting significantly slower rates of cyclization compared to the former (28). The slower rate of ProcM cyclization catalysis is juxtaposed with its ability to cyclize a plethora of natural substrates, which may be incompatible with high catalytic efficiency.

Inherent to the discussion of the enzymatically formed ring topologies are the stereochemistries of the α-carbon of the (Me)Lan rings (Figure 1.3C). Since dehydration and cyclization only involves the Ser or Thr sidechain and the thiol of Cys, the α-carbon of Cys remains untouched. Thus, two stereochemical outcomes are possible in the case of Lan, the 2*S*,6*R* isomer designated as DL-Lan and the 2*R*,6*R* isomer designated as LL-Lan. The additional methyl group at the former β-carbon of Thr also yields two isomers for MeLan, 2*S*,3*S*,6*R* for DL-MeLan and 2*R*,3*R*,6*R* for LL-MeLan (Figure 1.3C). The MeLan isomers 2*S*,3*R*,6*R* and 2*R*,3*S*,6*R* have not been found in lanthipeptides to date. Contrary to the DL-(Me)Lan configuration expected for all lanthipeptides, the thioether rings of the two peptides of the cytolysins were observed to have LL stereochemistry. Based on these observations as well as the sequence of the peptides it was hypothesized that the Dhx-Dhx-Xxx-Xxx-Cys motif was associated with the LL stereochemical outcome and was corroborated upon analysis of peptides bearing said motif (8, 29, 30).

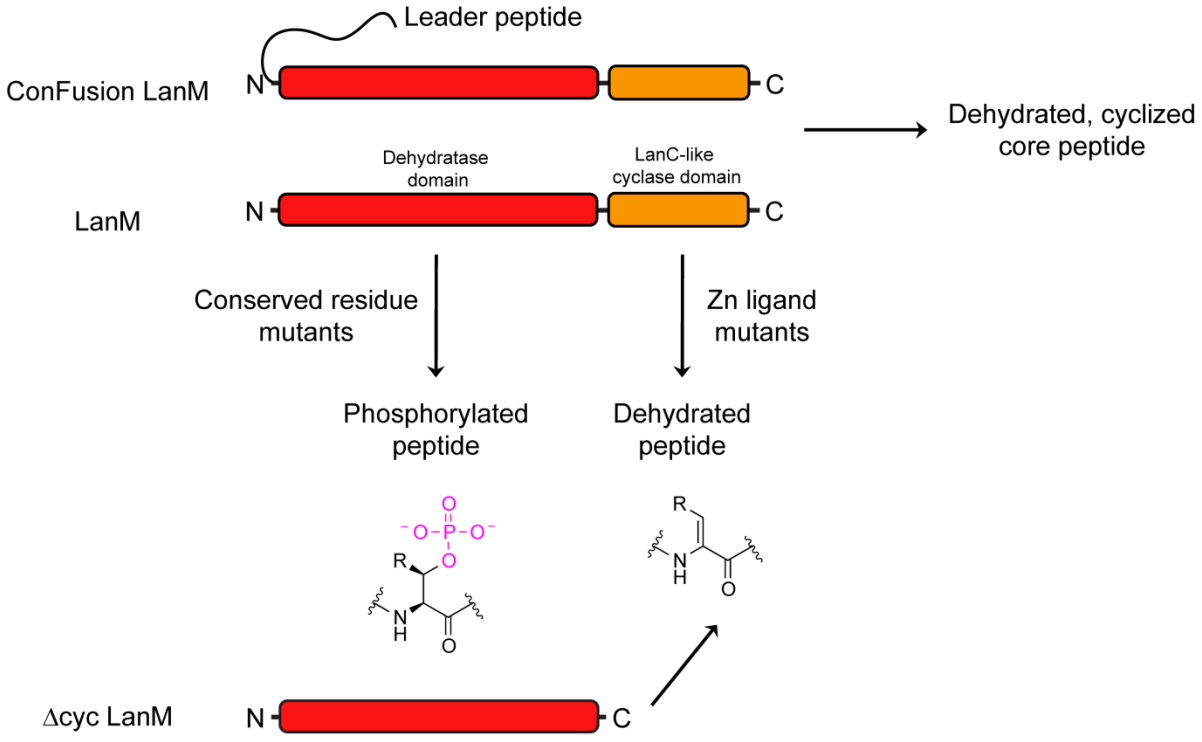

Figure 1.4. LanM bioengineering. Fusing the leader peptide to its cognate synthetase yields the ConFusion enzyme (4). Mutation of conserved residues within substrate-tolerant LanM proteins results in monofunctional enzymes capable of phosphorylating or dehydrating a variety of substrate peptides. Peptides containing dehydro amino acids can also be produced by using a truncated LanM enzyme lacking the cyclization domain (Δcyc LanM).

In addition to the strides made in understanding canonical LanM catalysis, non-canonical LanM enzymes have emerged and been the subject of investigation. In these cases, the dehydratase domain of the LanM can be identified but conserved residues in the cyclization domain are missing. Intriguingly, the gene clusters encoding for these non-canonical LanMs also often encode for reductases, which were demonstrated to reduce select dehydro amino acids installed by the cognate LanM protein into D-amino acids (8, 31, 32). Understandably these bioinformatically distinguished LanM proteins have yielded appreciation for the diversity of genetic contexts and corresponding biosynthetic roles that these enzymes assume, including in non-lantibiotic gene clusters (33).

1.3 In vivo and in vitro production of lanthipeptides

The approaches thus far reported for the production of lanthipeptides can be divided into two broad categories, in vivo or in vitro. The former can be further subdivided into strategies using the native organism or a heterologous host, and the latter into enzymatic or synthetic approaches (Figure 1.5). Reports of the development of genetic tools to activate gene clusters of interest followed by screening are limited, perhaps due to the genetic diversity of organisms that encode lanthipeptide gene clusters (34). In contrast, expression of gene clusters in heterologous hosts or reconstitution of enzyme activities in vitro has enjoyed widespread application. Complementary synthetic strategies have also been developed and have enabled access to additional chemical diversity. Most of the discussed strategies require leader peptide removal, which is discussed in further detail in section 1.4.

The small sizes of lanthipeptide biosynthetic gene clusters facilitates their amplification and cloning into plasmids for heterologous expression. Production of individual peptides and proteins is undertaken using established methods for protein production. In order to access the modified peptide, the peptide and protein genes are simultaneously expressed in the host. Subsequent purification of the peptide is accomplished using a combination of affinity and high performance liquid chromatographies (35). Due to the relative brevity of most lanthipeptide biosynthetic pathways (two to three enzymes for installation of the thioether rings), careful verification and optimization of enzyme expression is feasible. Heterologous production of lanthipeptides has been demonstrated only in the cases of class I and II systems, although expression of unmodified peptides and biosynthetic proteins is regularly performed for

lanthipeptides from all four classes (10). Cases of insoluble enzymes and/or substrates appear to be limited and can usually be overcome by the use of fusion proteins (36, 37). In addition, the heterologous expression of lanthipeptide systems is compatible with other protein engineering technologies such as unnatural amino acid incorporation, which has provided access to lanthipeptide substrates with non-proteinogenic functional groups (38-41). Expression of active class II synthetase enzymes in *E. coli* has proved especially facile, as reflected in the large and continuously increasing reports of reconstituted class II systems (42).

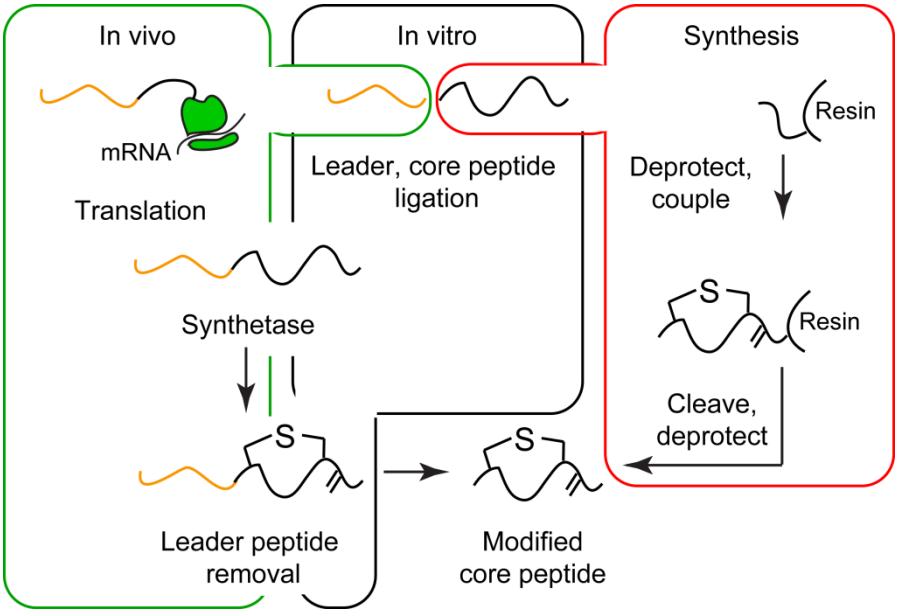

Figure 1.5. Strategies for lanthipeptide production. The leader peptide is indicated in orange, which can be produced by a heterologous host. Also, core peptides can be synthetically prepared and ligated to leader peptides in preparation for post-translational modification by a synthetase.

Although the heterologous production of peptides and proteins has been and continues to be fruitful, mechanistic information was limited through this approach. Thus, the in vitro reconstitution of lanthipeptide synthases was undertaken and has facilitated the study of not only the (Me)Lan-forming enzymes but also subsequent tailoring steps (12, 15, 43, 44). Detailed discussion of the resulting mechanistic insights can be found in section 1.3. Preparation for in vitro studies is procedurally similar to co-expression except that the LanA and synthase enzymes are expressed and purified separately (35). The demonstrated substrate tolerance of LanM

proteins, in part due to the modularity of the peptide scaffold, enabled in vitro biosynthetic modification of synthetic peptides. In these approaches synthetically prepared core peptide is ligated to a heterologously prepared leader peptide and exposed to the cognate synthase (Figure 1.5). Ligation is accomplished using expressed protein ligation or copper-catalyzed azide-alkyne cycloadditions (45, 46). The semi-synthetic core peptides can be prepared to serve as isotopically labelled substrates for mechanistic studies or post-translationally modified products that exhibit increased antimicrobial activity (47-49). Reliance on the leader peptide was more recently diminished by its incorporation into the synthase, thereby obviating the need for leader peptide ligation by producing a constitutively active enzyme (Figure 1.3) (4, 50). Thus far, in vitro preparation of lanthipeptide derivatives has heavily focused on the two class II synthetases LctM and ProcM. These studies highlight the impressive substrate tolerance and bioengineered capabilities of the class II biosynthetic machinery.

Lactocin S is the first lantibiotic to be synthesized on solid-support, and since then, the total synthesis on solid-support of several lantibiotics has been completed (51-54). Although procedures for coupling the peptide backbones are well established, installation of the thioether macrocycles in lanthipeptides presented a challenge. The strategy that was adopted involves preparation of synthetic (Me)Lan residues bearing orthogonal protecting groups on a set of amines and carboxylates, which allow for controlled and ordered macrocyclization and linear peptide elongation events. The selective coupling of one set of the (Me)Lan carboxylates and amines to the peptide chain is followed by deprotection of the second carboxylate of the (Me)Lan residue and coupling to form the thioether ring. Deprotection of the last amine allows for continued elongation of the peptide backbone (51-54). The ensuing discussions are focused on structure and activity relationships, with mechanistic observations discussed in section 1.3. Relatively minor structural changes to the thioether linkages in lacticin 3147 β result in analogues with retained but diminished antimicrobial activity. For example, a desmethyl variant of lacticin 3147 β, wherein the two MeLan rings are replaced with Lan rings, exhibited synergistic antimicrobial activity with its α component but 100-fold less than the natural compound (55). Progressive changes in the thioether rings of lacticin 3147 β by substitution of the sulfur atom with a pair of double bonded carbon atoms or oxygen atom resulted in compounds with diminished antimicrobial activities compared to the natural lantibiotic (56, 57). Similarly, synthetically accessed cytolysin and lacticin 481 diastereomers exhibited decreased

antimicrobial activity compared to their respective natural compounds, highlighting the importance of stereochemistry for activity (53, 54). However, some conservative perturbations are tolerated, for instance substitution of the A ring of lactocin S with a methylene group resulted in a variant with antimicrobial activity comparable to the natural lantibiotic (58).

1.4 Diversity, function, and removal of lanthipeptide leader peptides

The leader peptide is hypothesized to serve two functions during the biosynthetic maturation of lanthipeptides, the first is coordination of the biosynthetic events and the second is immunity. Succinct demonstration of the latter is provided by a qualitative decrease in the antimicrobial activity of full length, modified nisin as well as haloduracin α in combination with haloduracin β (59, 60). Illustrations of the former function are captured by the two proposed models of the role of the leader peptide during lanthipeptide post-translational modification: activation of synthase activity upon binding of leader peptide or ensnaring of an active enzyme subpopulation by the leader peptide (61). One observation that distinguishes between the two possibilities is that presentation of the LctA core peptide, in the absence of the LctA leader peptide to its corresponding synthetase LctM results in LctA core peptide modification (62). This outcome is incompatible with a strict requirement of the leader peptide for synthetase activation and is consistent with a limited portion of LctM enzymes being present in a conformationally active state. Although modification of the core peptide can be observed in the absence of leader peptide, core peptide alone is not sufficient for efficient modification. Indeed, subsequent iterations of employing LctM as a substrate tolerant catalyst fused the LctA leader peptide to the synthetase resulting in a constitutively active fusion enzyme termed ConFusion (Figure 1.4) (4). The improvement in LanA processing efficiency that the leader peptide provides is partially clarified by studies that quantify leader and core peptide binding to a synthetase. Specifically, the HalA2 leader peptide dissociation constant from HalM2 is significantly smaller than that of HalA2 core peptide. The contribution to binding that the HalA2 leader peptide offers is captured in the dissociation constants of full length HalA2 from HalM2 and the core peptide from ConFusion HalM2, which are similar to that of HalM2 for leader peptide alone (63). Perhaps the most revealing illustration of leader peptide importance is provided by studies into the processing of separated leader and core peptides. Systems in which the LctA leader peptide is separated from the core peptide exhibit differences in directionality and fidelity of modification

compared to the wild type system as well as an overall decrease in reaction rates, which explains the decreased efficiency of the ConFusion system for example (25).

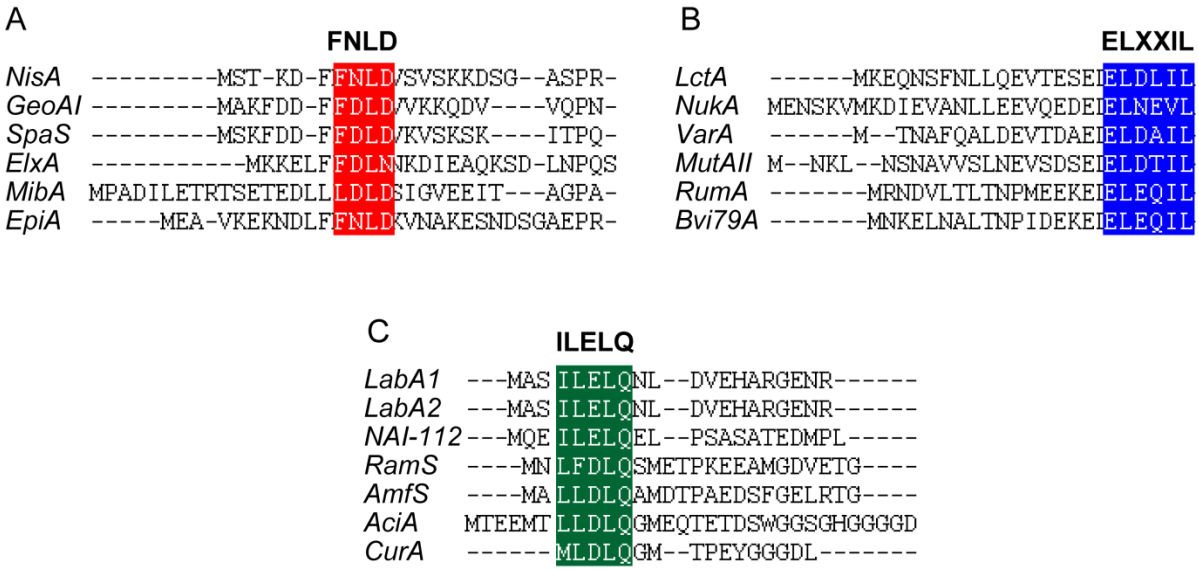

Figure 1.6. Leader peptide alignment and removal. Alignments of portions of the leader peptide of select members from each class of lanthipeptide. Conserved motifs are highlighted in red (A) for class I, blue (B) for class II and green (C) for class III. Reproduced from ref. (3).

Leader peptides from the different lanthipeptide classes exhibit diversity in their sequence and length, which parallels the diversity of the corresponding biosynthetic enzymes. In general, leader peptides from all lanthipeptide classes range from approximately 23 to 59 amino acids in length and have been proposed to adopt a helical conformation in select solvents and upon synthase binding (35, 64-66). In order to numerically discriminate between the leader and core peptides, the first, N-terminal residue of the core peptide is numbered "1" while the last C-terminal residue of the leader peptide is numbered "−1" with preceding N-terminal leader peptide residues bearing lower (i.e. more negative) numbers. Lanthipeptides from clusters in which more than two LanA substrates are encoded feature conserved leader peptides for all of the substrates (67-69), presumably due to convergence towards or duplication of a leader peptide sequence that efficiently interacts with its cognate synthase. In contrast, two-component lantibiotic systems, in which activity of an enzyme requires localization to a specific substrate, generally feature distinct leader peptides for the LanA1 and A2 precursor peptides. An exception is cytolysin for which the leader peptides of the two substrates have high homology. In some

cases, lanthipeptide leader peptides are bioinformatically annotated as nitrile hydratase-like (NHLP) or Nif11-like (68, 70), although the evolutionary underpinnings for this observation have not been elucidated. However other lanthipeptide leader peptides are not typically bioinformatically annotated, but do feature conserved amino acid sequence motifs that can be discerned in some cases (Figure 1.6) and include FNLD in class I leader peptides (71), GG/GA at the juncture between leader and core in class II leader peptides, and ILD/ELQ at the N-terminus of class III leader peptides (12, 66).

Although important for the post-translational modification of the precursor peptide, the leader peptide must be removed for manifestation of full activity (Figure 1.2A). In class I lanthipeptide systems, proteolysis of the leader peptide is completed by a LanP enzyme, which bears similarity to subtilisin-like serine proteases (44, 72). Leader peptide removal in class II systems is carried out by the bifunctional enzyme generally termed LanT, which also exports the final product into the extracellular space (73). LanT proteins are bioinformatically and structurally annotated as containing an N-terminal Cys protease domain, a transmembrane domain, and a C-terminal ATP-binding cassette (74). The apparent modularity could be translated into in vitro reconstitution of the Cys protease domain for the class II lantibiotic transporter LctT (75). Surprisingly, despite their encoding within genomes of Gram-positive bacteria, heterologous reconstitution of LanT activities in the Gram-negative heterologous host *E. coli* has proved successful in the case of the two-component lantibiotic lichenicidin (76), and as discussed in Chapter 2, may be generalizable. Preliminary support for this idea is presented by BovT150M, a fusion protein that combines the protease activity of the LanT BovT with the cognate synthetase BovM. BovT150M exhibited in vivo and in vitro activity in that it was able to modify and remove the leader peptide of its endogenous substrate BovA as well as a chimeric peptide consisting of the BovA leader and suicin core peptide (77). Lastly, leader peptide removal for class III lanthipeptides has only been characterized for the case of FlaP. This prolyl oligopeptidase exhibits some selectivity for dehydrated and cyclized peptide over unmodified, linear peptide (36). Finally, an appreciable number of lanthipeptides lack any of the aforementioned dedicated proteases encoded within their gene cluster, as exemplified by the microbisporicin gene cluster (78). In these cases it is hypothesized that other endogenous proteases encoded elsewhere in the genome affect leader peptide removal. Based on the limited

reports of leader peptide removal, the genetically encoded approaches to leader peptide removal could be as diverse as the strategies to dehydrate and cyclize the core peptides.

Intriguingly, additional N-terminal processing of the core peptide is observed in the cases of the two-component lantibiotics cytolysin, haloduracin, and lichenicidin (76, 79, 80). In the former case, haloduracin β isolated from the producing organism (*Bacillus halodurans* C-125) was observed to have a mass less than that calculated from proteolysis at the canonical GG/GA cleavage site. Furthermore, these additional amino acids from modified haloduracin β cut at the predicted site were removed upon exposure to *B. halodurans* C-125 culture supernatant. The responsible protease was not readily annotated by examination of the haloduracin gene cluster, since no additional proteases or protease domain containing enzymes (aside from HalT) were encoded in the gene cluster. In contrast, the lichenicidin gene cluster encodes both LicT and LicP for initial cleavage at the GG/GA motif and subsequent removal of a NDVNPE sequence, respectively (76, 81). Although the protein machinery for leader peptide removal is not fully conserved among the different classes of lanthipeptides, some dedicated lanthipeptide leader removal enzymes have demonstrated substrate tolerance. Aside from the NDVNPE recognition motif, LicP exhibits impressive substrate tolerance and was employed to efficiently proteolyze a variety of both linear and modified substrates (81).

Leader peptide removal in vitro is typically affected by commercial proteases, incorporation of a chemically cleavable moiety, or more recently by reconstitution of the activity of the endogenous protease. Although all of these strategies have been successfully employed, doing so requires optimization for each system. In the case of commercial proteases, the protease recognition site is usually not present in the core peptide such that the leader peptide is completely removed from the core peptide. However, incorporation of the commercial protease recognition site sometimes disrupts LanA processing, resulting in incomplete peptide modification (71, 82, 83). Disruption to processing can sometimes be minimized by incorporating non-amino acid functionalities between the leader and core peptides such as a base-labile ester linkage or a photocleavable aryl moiety (39, 84). Indeed, both approaches were executed to produce lacticin 481, but the requirement of unnatural amino acid incorporation or synthesis and ligation of the leader and core peptides can lower yields. In contrast, the alternative approach of reconstituting the activity of the LanT protease domain has only been demonstrated in three cases (32, 75, 77) and efforts to reconstitute the activities of protease domain truncants

of LanT proteins have produced inconsistent results. Furthermore, the lack of an encoded leader peptide removal enzyme in many lanthipeptide gene clusters precludes reconstitution of their activity. Conceivably, a suite of substrate-tolerant lanthipeptide proteases could provide the benefits of robust, traceless leader peptide removal without perturbation of the LanA sequence.

Efforts to identify leader peptide contacts with the lanthipeptide biosynthetic proteins were hampered by lack of synthase structures and a bioinformatically discernable domain. However, the co-crystallization of the leader peptide of NisA with NisB revealed interaction of the former with a winged helix-turn-helix motif structure within the latter (12). Notably, analogous structures, designated the RiPP precursor peptide recognition element (RRE), were identified from a collection of RiPP post-translational modification enzymes which included NisB (85). Additionally, the identified RRE appears distributed in over half of the representative RiPP gene clusters surveyed, further implicating leader peptide recognition during biosynthesis in RiPPs other than lanthipeptides. Despite repeated illustration of leader peptide importance, several lanthipeptide biosynthetic enzymes that introduce diversity beyond thioether crosslinks function in the absence of leader peptide (44, 86-88). This observation prompts caution against overstating the role of the leader peptide for biosynthesis, but does not detract from its established functions.

1.5 Lipid targets of lanthipeptides

Although lanthipeptides exhibit a variety of activities, definitive targets have been proposed only in the case of lantibiotics. Lipid II has been demonstrated and extrapolated as a target for most lantibiotics, while phosphatidylethanolamine is the proposed target of cinnamycin-like compounds (89, 90). Although extensive mutagenesis studies have been carried out on lantibiotics (91), fewer studies on the secondary structure of the peptides have been conducted. Furthermore, there have been only two reported structures of lanthipeptides in complex with their corresponding lipid target, which are discussed in more detail below. The involvement of secondary structure in target binding gained from the two available peptide-lipid complex structures and the roles of specific amino acids in these contexts bear some implications for the discovery of new lanthipeptides.

Many class I lantibiotics characterized to date exert their antimicrobial activity by binding the cell wall intermediate lipid II, and the low occurrence of resistance against

lantibiotics has been ascribed to their targeting this late stage cell wall biosynthesis intermediate (92). An NMR structure of the complex between nisin and lipid II revealed encapsulation of the pyrophosphate group of lipid II within a cage-like structure formed by the A and B rings of nisin (93) while the remaining hinge and C-terminal portion of the molecule are implicated in membrane insertion and pore formation. Although the hinge region does not directly contact lipid II, its functional importance was evidenced by mutagenesis of this region, which resulted in peptides with reduced antimicrobial activity relative to wild type nisin (94). Interestingly, biosynthesis of the lipid II binding nisin A/B ring is exclusively carried out by class I enzymes (34).

In addition to the lipid II binding nisin A/B ring motif, many lantibiotics biosynthesized by class II (LanM) enzymes feature another hypothesized lipid II binding motif found in the exemplary lantibiotic mersacidin. A six-membered thioether ring, usually the third ring of these peptides, is implicated in interacting with lipid II (95). Although NMR investigation of mersacidin in the presence of lipid II did not reveal direct contacts, significant conformational changes in the former were observed upon addition of the latter (96). Within the proposed lipid II binding motif of mersacidin is an indispensable Glu residue, which upon mutation results in a loss of activity (97). The proposed lipid II binding portion of mersacidin is also observed in class II, two-component lantibiotics. In these cases, the mersacidin-like lipid II binding portion is found in the α peptide, which presents a relatively more globular structure compared to the corresponding β peptides (Figure 1.2B). The model of α peptide-mediated recognition and binding to lipid II followed by β peptide-mediated pore-formation is supported by sequential addition and washing studies of lacticin 3147 and observed inhibition of transglycosylation and membrane disruption that haloduracin causes (60, 98-100). Lastly, the NMR structure of the lantibiotic cinnamycin in complex with lysophosphatidylethanolamine (PE) also highlights the importance of the (Me)Lan rings (101). Specifically, the ethanolamine headgroup of PE was nestled in a pocket formed by the B and C rings of the highly crosslinked and globular cinnamycin structure (Figure 1.2B). Paralleling the biosynthesis of the A/B ring motif of class I lantibiotics, the installation of the mersacidin-like lipid II binding motif appears to be monopolized by class II synthetases. Genome guided discovery (using either the LanA or M sequences) have routinely produced two-component lantibiotics featuring the mersacidin-like lipid II binding motif (42).

Lanthipeptides that do not exhibit antimicrobial activity do exhibit other activities including morphogenetic (102), antiallodynic (103), antiviral (7), and antifungal (104). However, definitive evidence for the targets of these peptides has not been presented. In addition, the continued discovery of new lanthipeptides which bear few sequence similarities to already characterized peptides presents a scenario in which attempts to extrapolate activity based on primary structure alone are difficult (31, 32). Lastly, the activities (physiological and otherwise) for a handful of other lanthipeptides (prochlorosins and venezuelins) remain unknown (37, 67). Given the characterized and hypothesized activities of lanthipeptides, it is tempting to speculate that the targets awaiting elucidation are cell wall or membrane associated components. However, in the absence of thorough screening efforts, the targets and cellular effects of the compounds will remain unknown. Although amino acid sequence conservation provides some reassurance for and indication of biological activity, the continued application of genome mining in pursuit of unusual lanthipeptide scaffolds leaves in its wake the non-trivial tasks of screening for activity and target identification.

1.6 In silico lanthipeptide discovery

Preliminary genome mining efforts for new lanthipeptides relied on manual querying of genomic databases with a sequence of interest (105). The two commonly surveyed databases are maintained by NCBI and the Department of Energy (Joint Genome Institute). Although the former boasts an exhaustive collection of genomes and encompasses the sequences found in Integrated Microbial Genomes & Microbiomes (IMG/M), the latter provides a more straightforward means of accessing and assessing potential gene clusters. Following the input of a lanthipeptide protein or peptide sequence of interest, the outputs of each program were manually surveyed and curated for "unusual" features in an iterative manner (the total or curated outputs were used as new query sequences) (106, 107). This approach is reliant on the user's assessment and thus depended on the user's familiarity with lanthipeptide literature. The ensuing sections highlight notable features of bioinformatics programs, challenges associated with genome mining operations, and projected trends for the resultant data. Several recent reviews offer comprehensive cataloging and discussion of available software (106, 108).

Due to the relatively small size of the open reading frames that often encode for RiPP precursor peptides, several recent programs offer improved sensitivity for their detection (109).

Alternative to the protein query strategy, specific organism genomes can be queried for a variety of natural product gene clusters by employing programs such as the antibiotics and Secondary Metabolites Analysis Shell (antiSMASH), which provides a pan-natural product assessment of the biosynthetic potential encoded in a particular genome (110). In conjunction with the influx of information, effective appreciation of the data necessitates an intuitive graphical user interface that allows for meaningful and customizable visualization of the large amounts of information. In contrast to the list-based output offered by NCBI's BLAST and antiSMASH, Cytoscape provides a network visualization that can be rearranged to reflect user-specified parameters such as the numerical quantification of similarity (111). Integration of output sequence information and attendant information is exemplified by output from the Enzyme Function Initiative (EFI), although this platform is not geared towards natural products per se (112, 113). A more relevant illustration is the arrangement of analyzed sequences into gene cluster families, from which a thorough overview of quantifiable correlations for each gene cluster can be gained (114).

Despite advances in computational applications which brought about newfound, systematic appreciation for the diversity of lanthipeptides and RiPPs in general (115), advancements in discovery continue to rely on rate limiting wet lab empiricism. Furthermore, the current functionality offered by existing in silico tools, with respect to new structure elucidation, can be achieved with additional time and application of preexisting methods. For example, the structural prediction capabilities of antiSMASH are currently limited, but continued support and expansion should improve accuracy (116). The current state of these programs does not detract from the systematic documentation they have provided or from their future utility. Lastly, the current carefully curated storehouses of published information do not enable facile, downloadable access to sequences and coupling to previously published data. The lack of centralized repository and standardization of information storage rendered transmission of gene cluster information suboptimal.

The repository Minimal Information about a Biosynthetic Gene cluster (MIBiG) hints at future trends towards a single platform that integrates the outputs of all of the platforms discussed above while requiring a single input. Certainly the call for standardization of nomenclature as well as the results of genome mining efforts reflect the popularity of and dedication to natural products research (61, 117). MIBiG attempts to address standardization, but currently relies on voluntary inputs, which may not be sufficiently compelling incentive for

participation. Providing access to the sequences in a standard format that is linked to relevant, published literature would allow users to build, integrate, and customize networks that address their research priorities. Standardizing data will not only facilitate gene cluster dissemination, but also indirectly refine a user's ability to generate structural hypotheses. An inherent difficulty in the discussion of these in silico tools is their structural prediction utility in the case of gene clusters that bear little similarity to characterized systems, at minimum they provide a means to assess perceived structural novelty. It is also foreseeable that the categorized data could feed future forays into automated synthetic biology.

The subsequent chapters describe my genome-guided contributions to the discovery of new lanthipeptides from a variety of microorganisms. Rather than the established method of screening physical samples for biological activities, a genomics approach was adopted, in which gene clusters of interest were identified *in silico* and the products elicited and characterized (118). The contents of this thesis relate examples of the latter approach. Specifically, Chapter 2 describes the *in silico* identification and characterization of a collection of lanthipeptides that are produced by an example of a combinatorial biosynthetic system. Similarly, Chapter 3 presents an example of employing genetic information to guide rapid identification of lanthipeptides from *Actinobacteria*. Chapter 4 details ongoing efforts to characterize new peptides from a combination of RiPP genetic heritages. In this vein, Chapter 5 discusses initial efforts to characterize the mechanism of catalysis of small LanB (sLanB) enzymes as well as the final product from a sLanB-rich gene cluster.

1.7 References

1. Newman, D.J., and Cragg, G.M. (2016). Natural products as sources of new drugs from 1981 to 2014. *J. Nat.Prod.* 79, 629-661.

2. Carlson, E.E. (2010). Natural products as chemical probes. *ACS Chem. Biol.* 5, 639-653.

3. Ortega, M.A. (2015). Biochemical characterization of enzymes involved in the post-translational modification of lantibiotics. *PhD Dissertation, University of Illinois at Urbana-Champaign.*

4. Oman, T.J., Knerr, P.J., Bindman, N.A., Velásquez, J.E., and van der Donk, W.A. (2012). An engineered lantibiotic synthetase that does not require a leader peptide on its substrate. *J. Am. Chem. Soc.* 134, 6952-6955.

5. Craik, D.J., Fairlie, D.P., Liras, S., and Price, D. (2013). The future of peptide-based drugs. *Chem. Biol. Drug Des.* 81, 136-147.

6. Arnison, P.G., Bibb, M.J., Bierbaum, G., Bowers, A.A., Bugni, T.S., Bulaj, G., Camarero, J.A., Campopiano, D.J., Challis, G.L., Clardy, J., Cotter, P.D., Craik, D.J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P.C., Entian, K.D., Fischbach, M.A., Garavelli, J.S., Göransson, U., Gruber, C.W., Haft, D.H., Hemscheidt, T.K., Hertweck, C., Hill, C., Horswill, A.R., Jaspars, M., Kelly, W.L., Klinman, J.P., Kuipers, O.P., Link, A.J., Liu, W., Marahiel, M.A., Mitchell, D.A., Moll, G.N., Moore, B.S., Müller, R., Nair, S.K., Nes, I.F., Norris, G.E., Olivera, B.M., Onaka, H., Patchett, M.L., Piel, J., Reaney, M.J.T., Rebuffat, S., Ross, R.P., Sahl, H.G., Schmidt, E.W., Selsted, M.E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Süssmuth, R.D., Tagg, J.R., Tang, G.L., Truman, A.W., Vederas, J.C., Walsh, C.T., Walton, J.D., Wenzel, S.C., Willey, J.M., and van der Donk, W.A. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* 30, 108-160.

7. Férir, G., Petrova, M.I., Andrei, G., Huskens, D., Hoorelbeke, B., Snoeck, R., Vanderleyden, J., Balzarini, J., Bartoschek, S., Brönstrup, M., Süssmuth, R.D., and Schols, D. (2013). The lantibiotic peptide labyrinthopeptin A1 demonstrates broad anti-HIV and anti-HSV activity with potential for microbicidal applications. *PLoS One.* 8, e64010.

8. Lohans, C.T., Li, J.L., and Vederas, J.C. (2014). Structure and biosynthesis of carnolysin, a homologue of enterococcal cytolysin with D-amino acids. *J. Am. Chem. Soc.* 136, 13150-13153.

9. Baindara, P., Chaudhry, V., Mittal, G., Liao, L.M., Matos, C.O., Khatri, N., Franco, O.L., Patil, P.B., and Korpole, S. (2015). Characterization of the antimicrobial peptide penisin, a class Ia novel lantibiotic from *Paenibacillus* sp. strain A3. *Antimicrob. Agents Chemother.* 60, 580-591.

10. Knerr, P.J., and van der Donk, W.A. (2012). Discovery, biosynthesis, and engineering of lantipeptides. *Annu. Rev. Biochem.* 81, 479-505.

11. Garg, N., Salazar-Ocampo, L.M., and van der Donk, W.A. (2013). In vitro activity of the nisin dehydratase NisB. *Proc. Natl. Acad. Sci. U.S.A.* 110, 7258-7263.

12. Ortega, M.A., Hao, Y., Zhang, Q., Walker, M.C., van der Donk, W.A., and Nair, S.K. (2015). Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature.* 517, 509-512.

13. Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., and Bateman, A. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279-285.

14. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410.

15. Xie, L., Miller, L.M., Chatterjee, C., Averin, O., Kelleher, N.L., and van der Donk, W.A. (2004). Lacticin 481: in vitro reconstitution of lantibiotic synthetase activity. *Science.* 303, 679-681.

16. You, Y.O., and van der Donk, W.A. (2007). Mechanistic investigations of the dehydration reaction of lacticin 481 synthetase using site-directed mutagenesis. *Biochemistry.* 46, 5991-6000.

17. Paul, M., Patton, G.C., and van der Donk, W.A. (2007). Mutants of the zinc ligands of lacticin 481 synthetase retain dehydration activity but have impaired cyclization activity. *Biochemistry.* 46, 6268-6276.

18. Dong, S.H., Tang, W., Lukk, T., Yu, Y., Nair, S.K., and van der Donk, W.A. (2015). The enterococcal cytolysin synthetase has an unanticipated lipid kinase fold. *elife.* 4, e07607.

19. You, Y.O., Levengood, M.R., Ihnken, L.A.F., Knowlton, A.K., and van der Donk, W.A. (2009). Lacticin 481 synthetase as a general serine/threonine kinase. *ACS Chem. Biol.* 4, 379-385.

20. Thibodeaux, G.N., and van der Donk, W.A. (2012). An engineered lantipeptide synthetase serves as a general leader peptide-dependent kinase. *Chem. Commun.* 48, 10615-10617.

21. Ma, H., Gao, Y., Zhao, F., and Zhong, J. (2015). Individual catalytic activity of two functional domains of bovicin HJ50 synthase BovM. *Wei Sheng Wu Xue Bao.* 55, 50-58.

22. Shimafuji, C., Noguchi, M., Nishie, M., Nagao, J., Shioya, K., Zendo, T., Nakayama, J., and Sonomoto, K. (2015). In vitro catalytic activity of N-terminal and C-terminal domains in NukM, the post-translational modification enzyme of nukacin ISK-1. *J. Biosci. Bioeng.* 120, 624-629.

23. Yu, Y., Mukherjee, S., and van der Donk, W.A. (2015). Product formation by the promiscuous lanthipeptide synthetase ProcM is under kinetic control. *J. Am. Chem. Soc.* 137, 5140-5148.

24. Tang, W., Thibodeaux, G.N., and van der Donk, W.A. (2016). The enterococcal cytolysin synthetase coevolves with substrate for stereoselective lanthionine synthesis. *ACS Chem. Biol.* 11, 2438-2446.

25. Thibodeaux, C.J., Wagoner, J., Yu, Y., and van der Donk, W.A. (2016). Leader peptide establishes dehydration order, promotes efficiency, and ensures fidelity during lacticin 481 biosynthesis. *J. Am. Chem. Soc.* 138, 6436-6444.

26. Repka, L.M., Chekan, J.R., Nair, S.K., and van der Donk, W.A. Mechanistic understanding of lanthipeptide biosynthetic enzymes. In press.

27. Lee, M.V., Ihnken, L.A., You, Y.O., McClerren, A.L., van der Donk, W.A., and Kelleher, N.L. (2009). Distributive and directional behavior of lantibiotic synthetases revealed by high-resolution tandem mass spectrometry. *J. Am. Chem. Soc.* 131, 12258-12264.

28. Thibodeaux, C.J., Ha, T., and van der Donk, W.A. (2014). A price to pay for relaxed substrate specificity: a comparative kinetic analysis of the class II lanthipeptide synthetases ProcM and HalM2. *J. Am. Chem. Soc.* 136, 17513-17529.

29. Tang, W., and van der Donk, W.A. (2013). The sequence of the enterococcal cytolysin imparts unusual lanthionine stereochemistry. *Nat. Chem. Biol.* 9, 157-159.

30. Garg, N., Goto, Y., Chen, T., and van der Donk, W.A. (2016). Characterization of the stereochemical configuration of lanthionines formed by the lanthipeptide synthetase GeoM. *Biopolymers.* 106, 834-842.

31. Yang, X., and van der Donk, W.A. (2015). Post-translational introduction of D-alanine into ribosomally synthesized peptides by the dehydroalanine reductase NpnJ. *J. Am. Chem. Soc.* 137, 12426-12429.

32. Huo, L., and van der Donk, W.A. (2016). Discovery and characterization of bicereucin, an unusual D-amino acid-containing mixed two-component lantibiotic. *J. Am. Chem. Soc.* 138, 5254-5257.

33. Freeman, M.F., Gurgui, C., Helf, M.J., Morinaka, B.I., Uria, A.R., Oldham, N.J., Sahl, H.G., Matsunaga, S., and Piel, J. (2012). Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. *Science.* 338, 387-390.

34. Zhang, Q., Yu, Y., Velásquez, J.E., and van der Donk, W.A. (2012). Evolution of lanthipeptide synthetases. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18361-18366.

35. Li, B., Cooper, L.E., and van der Donk, W.A. (2009). Chapter 21. In vitro studies of lantibiotic biosynthesis. *Methods Enzymol.* 458, 533-558.

36. Völler, G.H., Krawczyk, B., Ensle, P., and Süssmuth, R.D. (2013). Involvement and unusual substrate specificity of a prolyl oligopeptidase in class III lanthipeptide maturation. *J. Am. Chem. Soc.* 135, 7426-7429.

37. Goto, Y., Li, B., Claesen, J., Shi, Y.X., Bibb, M.J., and van der Donk, W.A. (2010). Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. *PLoS Biol.* 8, e1000339.

38. Oldach, F., Al Toma, R., Kuthning, A., Caetano, T., Mendo, S., Budisa, N., and Süssmuth, R.D. (2012). Congeneric lantibiotics from ribosomal in vivo peptide synthesis with noncanonical amino acids. *Angew. Chem. Int. Ed.* 51, 415-418.

39. Bindman, N.A., Bobeica, S.C., Liu, W.R., and van der Donk, W.A. (2015). Facile removal of leader peptides from lanthipeptides by incorporation of a hydroxy acid. *J. Am. Chem. Soc.* 137, 6975-6978.

40. Zhou, L., Shao, J.F., Li, Q., van Heel, A.J., de Vries, M.P., Broos, J., and Kuipers, O.P. (2016). Incorporation of tryptophan analogues into the lantibiotic nisin. *Amino Acids.* 48, 1309-1318.

41. Shi, Y., Yang, X., Garg, N., and van der Donk, W.A. (2011). Production of lantipeptides in *Escherichia coli*. *J. Am. Chem. Soc.* 133, 2338-2341.

42. Dischinger, J., Basi Chipalu, S., and Bierbaum, G. (2014). Lantibiotics: promising candidates for future applications in health care. *Int. J. Med. Microbiol.* 304, 51-62.

43. Li, B., Yu, J.P., Brunzelle, J.S., Moll, G.N., van der Donk, W.A., and Nair, S.K. (2006). Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis. *Science.* 311, 1464-1467.

44. Ortega, M.A., Velásquez, J.E., Garg, N., Zhang, Q., Joyce, R.E., Nair, S.K., and van der Donk, W.A. (2014). Substrate specificity of the lanthipeptide peptidase ElxP and the oxidoreductase ElxO. *ACS Chem. Biol.* 9, 1718-1725.

45. Zhang, X.G., and van der Donk, W.A. (2009). Using expressed protein ligation to probe the substrate specificity of lantibiotic synthetases. *Methods Enzymol.* 462, 117-134.

46. Levengood, M.R., Kerwood, C.C., Chatterjee, C., and van der Donk, W.A. (2009). Investigation of the Substrate Specificity of Lacticin 481 Synthetase by Using Nonproteinogenic Amino Acids. *ChemBioChem.* 10, 911-919.

47. Mukherjee, S., and van der Donk, W.A. (2014). Mechanistic studies on the substrate-tolerant lanthipeptide synthetase ProcM. *J. Am. Chem. Soc.* 136, 10450-10459.

48. Levengood, M.R., Knerr, P.J., Oman, T.J., and van der Donk, W.A. (2009). In vitro mutasynthesis of lantibiotic analogues containing nonproteinogenic amino acids. *J. Am. Chem. Soc.* 131, 12024-12025.

49. Krawczyk, B., Ensle, P., Müller, W.M., and Süssmuth, R.D. (2012). Deuterium labeled peptides give insights into the directionality of class III lantibiotic synthetase LabKC. *J. Am. Chem. Soc.* 134, 9922-9925.

50. Knerr, P.J., Oman, T.J., Garcia De Gonzalo, C.V., Lupoli, T.J., Walker, S., and van der Donk, W.A. (2012). Non-proteinogenic amino acids in lacticin 481 analogues result in more potent inhibition of peptidoglycan transglycosylation. *ACS Chem. Biol.* 7, 1791-1795.

51. Ross, A.C., Liu, H., Pattabiraman, V.R., and Vederas, J.C. (2010). Synthesis of the lantibiotic lactocin S using peptide cyclizations on solid phase. *J. Am. Chem. Soc.* 132, 462-463.

52. Knerr, P.J., and van der Donk, W.A. (2012). Chemical synthesis and biological activity of analogues of the lantibiotic epilancin 15X. *J. Am. Chem. Soc.* 134, 7648-7651.

53. Knerr, P.J., and van der Donk, W.A. (2013). Chemical synthesis of the lantibiotic lacticin 481 reveals the importance of lanthionine stereochemistry. *J. Am. Chem. Soc.* 135, 7094-7097.

54. Mukherjee, S., Huo, L., Thibodeaux, G.N., and van der Donk, W.A. (2016). Synthesis and bioactivity of diastereomers of the virulence lanthipeptide cytolysin. *Org. Lett.* 18, 6188-6191.

55. Pattabiraman, V.R., McKinnie, S.M., and Vederas, J.C. (2008). Solid-supported synthesis and biological evaluation of the lantibiotic peptide bis(desmethyl) lacticin 3147 A2. *Angew. Chem. Int. Ed. Engl.* 47, 9472-9475.

56. Pattabiraman, V.R., Stymiest, J.L., Derksen, D.J., Martin, N.I., and Vederas, J.C. (2007). Multiple on-resin olefin metathesis to form ring-expanded analogues of the lantibiotic peptide, lacticin 3147 A2. *Org. Lett.* 9, 699-702.

57. Liu, H., Pattabiraman, V.R., and Vederas, J.C. (2009). Synthesis and biological activity of oxa-lacticin A2, a lantibiotic analogue with sulfur replaced by oxygen. *Org. Lett.* 11, 5574-5577.

58. Ross, A.C., McKinnie, S.M., and Vederas, J.C. (2012). The synthesis of active and stable diaminopimelate analogues of the lantibiotic peptide lactocin S. *J. Am. Chem. Soc.* 134, 2008-2011.

59. van der Meer, J.R., Polman, J., Beerthuyzen, M.M., Siezen, R.J., Kuipers, O.P., and De Vos, W.M. (1993). Characterization of the *Lactococcus lactis* nisin A operon genes nisP, encoding a subtilisin-like serine protease involved in precursor processing, and nisR, encoding a regulatory protein involved in nisin biosynthesis. *J. Bacteriol.* 175, 2578-2588.

60. Oman, T.J., Lupoli, T.J., Wang, T.S., Kahne, D., Walker, S., and van der Donk, W.A. (2011). Haloduracin alpha binds the peptidoglycan precursor lipid II with 2:1 stoichiometry. *J. Am. Chem. Soc.* 133, 17544-17547.

61. Yang, X., and van der Donk, W.A. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: new insights into the role of leader and core peptides during biosynthesis. *Chem. Eur. J.* 19, 7662-7677.

62. Levengood, M.R., Patton, G.C., and van der Donk, W.A. (2007). The leader peptide is not required for post-translational modification by lacticin 481 synthetase. *J. Am. Chem. Soc.* 129, 10314-10315.

63. Thibodeaux, G.N., McClerren, A.L., Ma, Y., Gancayco, M.R., and van der Donk, W.A. (2015). Synergistic binding of the leader and core peptides by the lantibiotic synthetase HalM2. *ACS Chem. Biol.* 10, 970-977.

64. Patton, G.C., Paul, M., Cooper, L.E., Chatterjee, C., and van der Donk, W.A. (2008). The importance of the leader sequence for directing lanthionine formation in lacticin 481. *Biochemistry.* 47, 7342-7351.

65. Nagao, J., Morinaga, Y., Islam, M.R., Asaduzzaman, S.M., Aso, Y., Nakayama, J., and Sonomoto, K. (2009). Mapping and identification of the region and secondary structure required for the maturation of the nukacin ISK-1 prepeptide. *Peptides.* 30, 1412-1420.

66. Müller, W.M., Ensle, P., Krawczyk, B., and Süssmuth, R.D. (2011). Leader peptide-directed processing of labyrinthopeptin A2 precursor peptide by the modifying enzyme LabKC. *Biochemistry.* 50, 8362-8373.

67. Li, B., Sher, D., Kelly, L., Shi, Y.X., Huang, K., Knerr, P.J., Joewono, I., Rusch, D., Chisholm, S.W., and van der Donk, W.A. (2010). Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10430-10435.

68. Zhang, Q., Yang, X., Wang, H., and van der Donk, W.A. (2014). High divergence of the precursor peptides in combinatorial lanthipeptide biosynthesis. *ACS Chem. Biol.* 9, 2686-2694.

69. Maricic, N., Anderson, E.S., Opipari, A.E., Yu, E.A., and Dawid, S. (2016). Characterization of a multipeptide lantibiotic locus in *Streptococcus pneumoniae*. *MBio.* 7, e01656-01615.

70. Haft, D.H., Basu, M.K., and Mitchell, D.A. (2010). Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family. *BMC Biol.* 8, 70.

71. Plat, A., Kluskens, L.D., Kuipers, A., Rink, R., and Moll, G.N. (2011). Requirements of the engineered leader peptide of nisin for inducing modification, export, and cleavage. *Appl. Environ. Microbiol.* 77, 604-611.

72. Siezen, R.J., Rollema, H.S., Kuipers, O.P., and de Vos, W.M. (1995). Homology modelling of the *Lactococcus lactis* leader peptidase NisP and its interaction with the precursor of the lantibiotic nisin. *Protein Eng.* 8, 117-125.

73. Håvarstein, L.S., Diep, D.B., and Nes, I.F. (1995). A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Mol. Microbiol.* 16, 229-240.

74. Lin, D.Y., Huang, S., and Chen, J. (2015). Crystal structures of a polypeptide processing and secretion transporter. *Nature.* 523, 425-430.

75. Ihnken, L.A.F., Chatterjee, C., and van der Donk, W.A. (2008). In vitro reconstitution and substrate specificity of a lantibiotic protease. *Biochemistry.* 47, 7352-7363.

76. Caetano, T., Krawczyk, J.M., Mösker, E., Süssmuth, R.D., and Mendo, S. (2011). Heterologous expression, biosynthesis, and mutagenesis of type II lantibiotics from *Bacillus licheniformis* in *Escherichia coli*. *Chem. Biol.* 18, 90-100.

77. Wang, J., Ge, X., Zhang, L., Teng, K., and Zhong, J. (2016). One-pot synthesis of class II lanthipeptide bovicin HJ50 via an engineered lanthipeptide synthetase. *Sci. Rep.* 6, 38630.

78. Foulston, L.C., and Bibb, M.J. (2010). Microbisporicin gene cluster reveals unusual features of lantibiotic biosynthesis in actinomycetes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13461-13466.

79. Cooper, L.E., McClerren, A.L., Chary, A., and van der Donk, W.A. (2008). Structure-activity relationship studies of the two-component lantibiotic haloduracin. *Chem. Biol.* 15, 1035-1045.

80. Booth, M.C., Bogie, C.P., Sahl, H.G., Siezen, R.J., Hatter, K.L., and Gilmore, M.S. (1996). Structural analysis and proteolytic activation of *Enterococcus faecalis* cytolysin, a novel lantibiotic. *Mol. Microbiol.* 21, 1175-1184.

81. Tang, W.X., Dong, S.H., Repka, L.M., He, C., Nair, S.K., and van der Donk, W.A. (2015). Applications of the class II lanthipeptide protease LicP for sequence-specific, traceless peptide bond cleavage. *Chem. Sci.* 6, 6270-6279.

82. Garg, N., Tang, W., Goto, Y., Nair, S.K., and van der Donk, W.A. (2012). Lantibiotics from *Geobacillus thermodenitrificans*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5241-5246.

83. Tang, W., and van der Donk, W.A. (2012). Structural characterization of four prochlorosins: a novel class of lantipeptides produced by planktonic marine cyanobacteria. *Biochemistry.* 51, 4271-4279.

84. Bindman, N., Merkx, R., Koehler, R., Herrman, N., and van der Donk, W.A. (2010). Photochemical cleavage of leader peptides. *Chem. Commun.* 46, 8935-8937.

85. Burkhart, B.J., Hudson, G.A., Dunbar, K.L., and Mitchell, D.A. (2015). A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nat. Chem. Biol.* 11, 564-570.

86. Kupke, T., Kempter, C., Jung, G., and Götz, F. (1995). Oxidative decarboxylation of peptides catalyzed by flavoprotein EpiD. Determination of substrate specificity using peptide libraries and neutral loss mass spectrometry. *J. Biol. Chem.* 270, 11282-11289.

87. Ökesli, A., Cooper, L.E., Fogle, E.J., and van der Donk, W.A. (2011). Nine post-translational modifications during the biosynthesis of cinnamycin. *J. Am. Chem. Soc.* 133, 13753-13760.

88. Ortega, M.A., Cogan, D.P., Mukherjee, S., Garg, N., Li, B., Thibodeaux, G.N., Maffioli, S., Donadio, S., Sosio, M., Escano, J., Smith, L., Nair, S.K., and van der Donk, W.A.

(2016). Two flavoenzymes catalyze the post-translational generation of 5-chlorotryptophan and 2-aminovinyl-cysteine during NAI-107 biosynthesis. *ACS Chem. Biol.*, 10.1021/acschembio.1026b01031.

89. Breukink, E., and de Kruijff, B. (2006). Lipid II as a target for antibiotics. *Nat. Rev. Drug Discovery.* 5, 321-332.

90. Zhao, M. (2011). Lantibiotics as probes for phosphatidylethanolamine. *Amino Acids.* 41, 1071-1079.

91. Ross, A.C., and Vederas, J.C. (2011). Fundamental functionality: recent developments in understanding the structure-activity relationships of lantibiotic peptides. *J. Antibiot.* 64, 27-34.

92. Breukink, E., Wiedemann, I., van Kraaij, C., Kuipers, O.P., Sahl, H.G., and de Kruijff, B. (1999). Use of the cell wall precursor lipid II by a pore-forming peptide antibiotic. *Science.* 286, 2361-2364.

93. Hsu, S.T., Breukink, E., Tischenko, E., Lutters, M.A., de Kruijff, B., Kaptein, R., Bonvin, A.M., and van Nuland, N.A. (2004). The nisin-lipid II complex reveals a pyrophosphate cage that provides a blueprint for novel antibiotics. *Nat. Struct. Mol. Biol.* 11, 963-967.

94. Wiedemann, I., Breukink, E., van Kraaij, C., Kuipers, O.P., Bierbaum, G., de Kruijff, B., and Sahl, H.G. (2001). Specific binding of nisin to the peptidoglycan precursor lipid II combines pore formation and inhibition of cell wall biosynthesis for potent antibiotic activity. *J. Biol. Chem.* 276, 1772-1779.

95. Dufour, A., Hindré, T., Haras, D., and Le Pennec, J.P. (2007). The biology of lantibiotics from the lacticin 481 group is coming of age. *FEMS Microbiol. Rev.* 31, 134-167.

96. Hsu, S.T., Breukink, E., Bierbaum, G., Sahl, H.G., de Kruijff, B., Kaptein, R., van Nuland, N.A., and Bonvin, A.M. (2003). NMR study of mersacidin and lipid II interaction in dodecylphosphocholine micelles. Conformational changes are a key to antimicrobial activity. *J. Biol. Chem.* 278, 13110-13117.

97. Szekat, C., Jack, R.W., Skutlarek, D., Farber, H., and Bierbaum, G. (2003). Construction of an expression system for site-directed mutagenesis of the lantibiotic mersacidin. *Appl. Environ. Microbiol.* 69, 3777-3783.

98. Oman, T.J., and van der Donk, W.A. (2009). Insights into the mode of action of the two-peptide lantibiotic haloduracin. *ACS Chem. Biol.* 4, 865-874.

99. Morgan, S.M., O'Connor, P.M., Cotter, P.D., Ross, R.P., and Hill, C. (2005). Sequential actions of the two component peptides of the lantibiotic lacticin 3147 explain its antimicrobial activity at nanomolar concentrations. *Antimicrob. Agents Chemother.* 49, 2606-2611.

100. Wiedemann, I., Bottiger, T., Bonelli, R.R., Wiese, A., Hagge, S.O., Gutsmann, T., Seydel, U., Deegan, L., Hill, C., Ross, P., and Sahl, H.G. (2006). The mode of action of the lantibiotic lacticin 3147-a complex mechanism involving specific interaction of two peptides and the cell wall precursor lipid II. *Mol. Microbiol.* 61, 285-296.

101. Hosoda, K., Ohya, M., Kohno, T., Maeda, T., Endo, S., and Wakamatsu, K. (1996). Structure determination of an immunopotentiator peptide, cinnamycin, complexed with lysophosphatidylethanolamine by 1H-NMR1. *J. Biochem.* 119, 226-230.

102. Willey, J.M., and Gaskell, A.A. (2011). Morphogenetic signaling molecules of the streptomycetes. *Chem. Rev.* 111, 174-187.

103. Meindl, K., Schmiederer, T., Schneider, K., Reicke, A., Butz, D., Keller, S., Gühring, H., Vértesy, L., Wink, J., Hoffmann, H., Brönstrup, M., Sheldrick, G.M., and Süssmuth, R.D. (2010). Labyrinthopeptins: a new class of carbacyclic lantibiotics. *Angew. Chem. Int. Ed.* 49, 1151-1154.

104. Mohr, K.I., Volz, C., Jansen, R., Wray, V., Hoffmann, J., Bernecker, S., Wink, J., Gerth, K., Stadler, M., and Müller, R. (2015). Pinensins: the first antifungal lantibiotics. *Angew. Chem. Int. Ed.* 54, 11254-11258.

105. Velásquez, J.E., and van der Donk, W.A. (2011). Genome mining for ribosomally synthesized natural products. *Curr. Opin. Chem. Biol.* 15, 11-21.

106. Ziemert, N., Alanjary, M., and Weber, T. (2016). The evolution of genome mining in microbes - a review. *Nat. Prod. Rep.* 33, 988-1005.

107. Medema, M.H., and Fischbach, M.A. (2015). Computational approaches to natural product discovery. *Nat. Chem. Biol.* 11, 639-648.

108. Hetrick, K.J., and van der Donk, W.A. Ribosomally synthesized and post-translationally modified peptide natural product discovery in the genomic era. *Curr. Opin. Chem. Biol.* Submitted.

109. Tietz, J.I., Schwalen, C.J., Patel, P.S., Maxon, T., Blair, P.M., Tai, H.C., Zakai, U.I., and Mitchell, D.A. A new genome mining tool redefines the lasso peptide biosynthetic landscape. *Nat. Chem. Biol.* In press.

110. Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., and Breitling, R. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 39, W339-346.

111. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498-2504.

112. Gerlt, J.A., Allen, K.N., Almo, S.C., Armstrong, R.N., Babbitt, P.C., Cronan, J.E., Dunaway-Mariano, D., Imker, H.J., Jacobson, M.P., Minor, W., Poulter, C.D., Raushel, F.M., Sali, A., Shoichet, B.K., and Sweedler, J.V. (2011). The Enzyme Function Initiative. *Biochemistry.* 50, 9950-9962.

113. Gerlt, J.A., Bouvier, J.T., Davidson, D.B., Imker, H.J., Sadkhin, B., Slater, D.R., and Whalen, K.L. (2015). Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta.* 1854, 1019-1037.

114. Doroghazi, J.R., Albright, J.C., Goering, A.W., Ju, K.S., Haines, R.R., Tchalukov, K.A., Labeda, D.P., Kelleher, N.L., and Metcalf, W.W. (2014). A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol.* 10, 963-968.

115. Skinnider, M.A., Johnston, C.W., Edgar, R.E., Dejong, C.A., Merwin, N.J., Rees, P.N., and Magarvey, N.A. (2016). Genomic charting of ribosomally synthesized natural product chemical space facilitates targeted mining. *Proc. Natl. Acad. Sci. U.S.A.* 113, E6343-E6351.

116. Blin, K., Kazempour, D., Wohlleben, W., and Weber, T. (2014). Improved lanthipeptide detection and prediction for antiSMASH. *PLoS One.* 9, e89420.

117. Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K., de Bruijn, I., Chooi, Y.H., Claesen, J., Coates, R.C., Cruz-Morales, P., Duddela, S., Düsterhus, S., Edwards, D.J., Fewer, D.P., Garg, N., Geiger, C., Gomez-Escribano, J.P., Greule, A., Hadjithomas, M., Haines, A.S., Helfrich, E.J.N., Hillwig, M.L., Ishida, K., Jones, A.C., Jones, C.S., Jungmann, K., Kegler, C., Kim, H.U., Kötter, P., Krug, D., Masschelein, J., Melnik, A.V., Mantovani, S.M., Monroe, E.A., Moore, M., Moss, N., Nützmann, H.W., Pan, G.H., Pati, A., Petras, D., Reen, F.J., Rosconi, F., Rui, Z., Tian, Z.H., Tobias, N.J., Tsunematsu, Y., Wiemann, P., Wyckoff, E., Yan, X.H., Yim, G., Yu, F.G., Xie, Y.C., Aigle, B., Apel, A.K., Balibar, C.J., Balskus, E.P., Barona-Gómez, F., Bechthold, A., Bode, H.B., Borriss, R., Brady, S.F., Brakhage, A.A., Caffrey, P., Cheng, Y.Q., Clardy, J., Cox, R.J., De Mot, R., Donadio, S., Donia, M.S., van der Donk, W.A., Dorrestein, P.C., Doyle, S., Driessen, A.J.M., Ehling-Schulz, M., Entian, K.D., Fischbach, M.A., Gerwick, L., Gerwick, W.H., Gross, H., Gust, B., Hertweck, C., Höfte, M., Jensen, S.E., Ju, J.H., Katz, L., Kaysser, L., Klassen, J.L., Keller, N.P., Kormanec, J., Kuipers, O.P., Kuzuyama, T., Kyrpides, N.C., Kwon, H.J., Lautru, S., Lavigne, R., Lee, C.Y., Linquan, B., Liu, X.Y., Liu, W., Luzhetskyy, A., Mahmud, T., Mast, Y., Méndez, C., Metsä-Ketelä, M., Micklefield, J., Mitchell, D.A., Moore, B.S., Moreira, L.M., Müller, R., Neilan, B.A., Nett, M., Nielsen, J., O'Gara, F., Oikawa, H., Osbourn, A., Osburne, M.S., Ostash, B., Payne, S.M., Pernodet, J.L., Petricek, M., Piel, J., Ploux, O., Raaijmakers, J.M., Salas, J.A., Schmitt, E.K., Scott, B., Seipke, R.F., Shen, B., Sherman, D.H., Sivonen, K., Smanski, M.J., Sosio, M., Stegmann, E., Süssmuth, R.D., Tahlan, K., Thomas, C.M., Tang, Y., Truman, A.W., Viaud, M., Walton, J.D., Walsh, C.T., Weber, T., van Wezel, G.P., Wilkinson, B., Willey, J.M., Wohlleben, W., Wright, G.D., Ziemert,

N., Zhang, C.S., Zotchev, S.B., Breitling, R., Takano, E., and Glockner, F.O. (2015). Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* 11, 625-631.

118. Tietz, J.I., and Mitchell, D.A. (2016). Using genomics for natural product structure elucidation. *Curr. Top. Med. Chem.* 16, 1645-1694.

Chapter 2. Production of lanthipeptides from the anaerobe *Ruminococcus flavefaciens* FD-1 in *Escherichia coli*

2.1 Introduction

The human microbiome has received much attention in recent years for its connection to human health (1-3). Increasingly, the mutualistic relationship between the microbiome and host has been demonstrated (4-7). One of the beneficial roles of the gut microbiota of a healthy individual is believed to be the resistance that is provided against colonization by pathogens (8). Similarly, the microbiota of livestock is critical for animal health, and a better understanding of the mechanisms that confer pathogen resistance in this setting is desired. Ruminant animals such as cattle and sheep have a symbiotic relationship with ruminal microorganisms that can degrade cellulose and/or hemicellulose (9, 10). Among the bacteria with cellulose activity in the rumen are the anaerobic ruminococci (11). Some *Ruminococcus* strains have been reported to produce bacteriocins (12-16), a class of ribosomally produced antimicrobial compounds that may be important for maintaining a niche in the competitive microbial environment of the rumen as well as in the human gastrointestinal tract (17). *Ruminococcus flavefaciens* FD-1 has a particularly high cellulolytic activity (18). It has not been reported to produce any antimicrobial peptides, but its genome was recently sequenced (19), providing a view of the genetic capability to produce such compounds. Here we show that its genome encodes an unusual group of lanthipeptides that are composed of four highly conserved copies of a peptide that likely binds lipid II and a diverse set of eight additional peptides, some of which act synergistically with the lipid II-binding peptide. The possible functional implications of the system in the context of the rumen environment are discussed in this Chapter.

As mentioned in Chapter 1, lanthipeptides that display antibacterial activity are called lantibiotics, and interest in these compounds has stemmed from their potent antibacterial activity and low propensity for the development of resistance (20). Although the prospects of RiPP production in anaerobic organisms has been highlighted (21), relatively few such compounds produced by the genus *Ruminococcus* have been reported thus far (14, 22, 23). We designated the lanthipeptide biosynthetic gene cluster in *R. flavefaciens* FD-1 *flv* (Figure 2.1A). Compared to known lanthipeptide biosynthetic gene clusters, the *flv* cluster possesses unusual characteristics that drew our attention. The cluster encodes twelve putative *lanA* genes and two

putative *lanM* genes (Figure 2.1A). The large number of substrates is reminiscent of the prochlorosin-like systems (24, 25) and the presence of two LanM synthetases resembles two-component lantibiotic systems such as lacticin 3147 and haloduracin (26, 27). As discussed in Chapter 1, most of the two-component lanthipeptides investigated to date consist of a lipid II-binding α-peptide and a β-peptide that acts synergistically to affect pore formation via a poorly defined mechanism (28-32). The gene clusters for two-component lantibiotics usually encode a single copy of an α and a β-peptide, and hence the twelve FlvA peptides encoded within the *flv* gene cluster are unprecedented. In light of these unusual characteristics, we set out to obtain the modified FlvA peptides. Systematic production of these peptides facilitated their structural characterization and bioactivity analysis.

2.2 Results

2.2.1 Bioinformatic analysis of the FlvA substrates

Inspection of the FlvA amino acid sequences revealed that four out of the twelve FlvA peptides possessed nearly identical amino acid sequences, with a single and two amino acid differences in the leader and core peptide, respectively (Figure 2.1B). Querying the non-redundant protein database from the NCBI using the BLAST (33) with the core peptides from this nearly identical set of FlvA peptides returned the LicA1 peptide sequence (Figure 2.1C) (28, 34-36). The modified LicA1 core peptide forms the α-peptide of the two-component lantibiotic lichenicidin. An alignment of the lichenicidin peptides and all of the FlvA peptides using the alignment program Clustal Omega (37) indicated that the latter could be roughly divided into a set of putative α and β precursor peptides (Figure 2.1C, D). The four nearly identical FlvA peptides (Figure 2.1B) make up the putative α-peptides and were designated FlvA1.x (where x = a-d), whereas the remaining eight FlvA peptides were hypothesized to be the corresponding β-peptides and were designated FlvA2.x (where x = a-h). Whereas the FlvA1.x core peptide amino acid sequences resemble that of lichenicidin α, the core peptides of FlvA2.x exhibited more limited sequence similarity with lichenicidin β, although some of the Ser, Thr, and Cys residues in the C-termini were similarly positioned (Figure 2.1D). In general, the core peptides of the FlvA peptides are longer than those of the previously characterized α- and β-peptides.

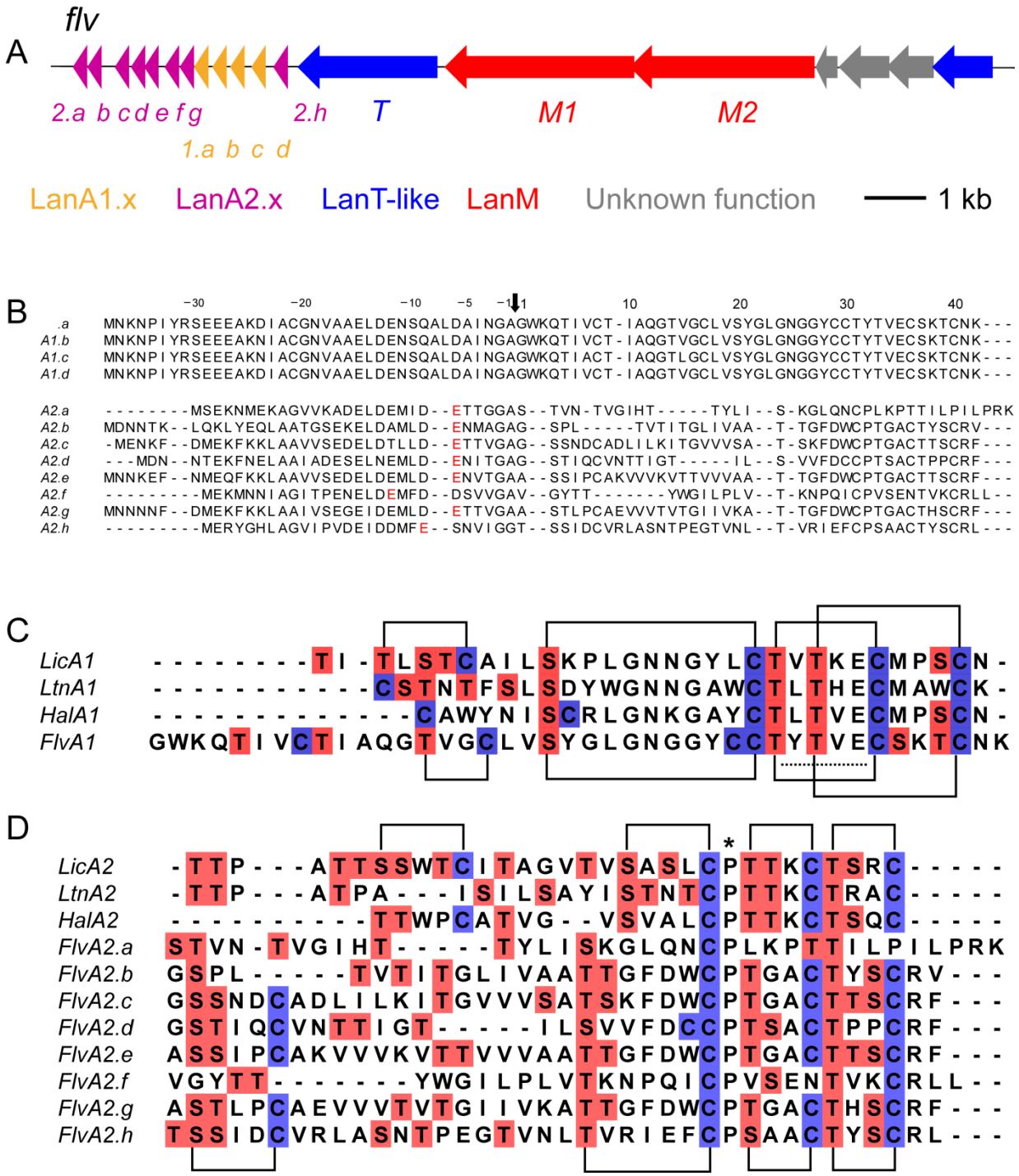

Figure 2.1. Comparison of the *flv* system with known two-component lantibiotics. All alignments were performed using Clustal Omega. (A) The *flv* biosynthetic cluster encoded within the genome of *R. flavefaciens* FD-1. (B) Alignment of the full length FlvA peptides. The Glu at which endoproteinase Glu-C cleaves in each of the modified FlvA2.x peptides is shown in red. An arrow indicates the predicted endogenous FlvT cleavage site in the FlvA

Figure 2.1. (cont.)

peptides. (C) Alignment of the FlvA1.a core peptide with previously reported α peptides of two-component lantibiotics. Dehydratable residues in the core peptides are highlighted in red and cysteine residues are highlighted in blue. The reported ring topology of lichenicidin is indicated above the sequence, and the proposed ring topology for the modified FlvA1 peptide is shown below the sequence. Hal denotes the locus for haloduracin, Ltn denotes the locus for lacticin 3147, and Lic the locus for lichenicidin. The conserved lipid II binding motif is indicated with a dotted line. (D) Alignment of the FlvA2.x peptides with reported β peptides of two-component lantibiotics. The fully conserved Pro is indicated with an asterisk.

Further comparison of the FlvA peptides with other two-component lantibiotics revealed additional shared and unusual features. The FlvA1.a peptide has Ser20 and Cys30 positioned to form a putative lipid II binding motif (Figure 2.1C, underlined) discussed in Section 1.5 that is found in the lantibiotics mersacidin, lacticin 481, and nukacin ISK-1 (38-41) and in the α-peptide of two-component lantibiotics (28, 29, 34, 42). However, compared to the other α-peptides, the FlvA1.x peptides contain more Cys residues in the core peptide (Figure 2.1C). The involvement of these Cys residues in thioether rings or disulfide bridges could endow the FlvA1.a peptide with increased stability, which may be important in the ruminal environment. Similar to the FlvA1.x peptides, many of the FlvA2.x core peptides (c-e, g, and h) also have additional Ser and Thr residues compared to other two-component β-peptides (Figure 2.1D). Notably, some of these residues form a putative Dhx-Dhx-Xxx-Xxx-Cys motif (where Dhx = Dha or Dhb and Xxx = any amino acid), which has been shown in some lanthipeptides to result in MeLan rings with LL stereochemistry (43-45). Corollary to the structural diversity of the FlvA2.x peptides would be a high substrate tolerance of the FlvM enzyme(s) responsible for installing the (Me)Lan rings.

2.2.2 Attempted detection of modified FlvA production in *R. flavefaciens* FD-1

Liquid cultures of anaerobically grown *R. flavefaciens* FD-1 were desalted and analyzed for the presence of FlvA peptides. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) analysis did not provide clear indication of masses corresponding to putative peptide products of the *flv* cluster. In alternative attempts to elicit peptide production, *R. flavefaciens* FD-1 and *Ruminococcus albus* 7 were co-cultured as well as grown in proximity to one another on solid media, an approach that has been successful for other

genera (46, 47). However, no new ions with m/z > 2000 could be observed by MALDI-TOF MS. These observations are representative of a commonly encountered hurdle in natural product research: the environmental triggers for expression of biosynthetic genes are generally poorly understood and often cannot be replicated in the laboratory (48, 49). The inability to detect potential products of the *flv* cluster prompted consideration of alternate strategies to produce and assess the products generated from the *flv* genes.

2.2.3 Heterologous production of modified FlvA peptides

To access the products of the Flv system, a validated heterologous production strategy using *Escherichia coli* was employed. This approach has been successfully applied to the production of several lantibiotics and involves the co-expression of a LanA peptide and cognate LanM enzyme in *E. coli*, which typically results in full post-translational modification of the peptide. The modified full length LanA peptide is then purified from *E. coli* and the leader peptide is removed in vitro using a variety of proteolytic strategies as discussed in Section 1.4 (36, 43, 50-56). The sequence similarity of FlvM1 with characterized two-component LanM1 enzymes is greater than the similarity of the latter enzymes with FlvM2. Thus, we hypothesized that FlvM1 was responsible for the modification of the putative FlvA1.x peptides and FlvM2 for the FlvA2.x peptides. To probe this hypothesis, we undertook pilot co-expression studies of an arbitrarily chosen peptide FlvA2.g with either FlvM1 or FlvM2.

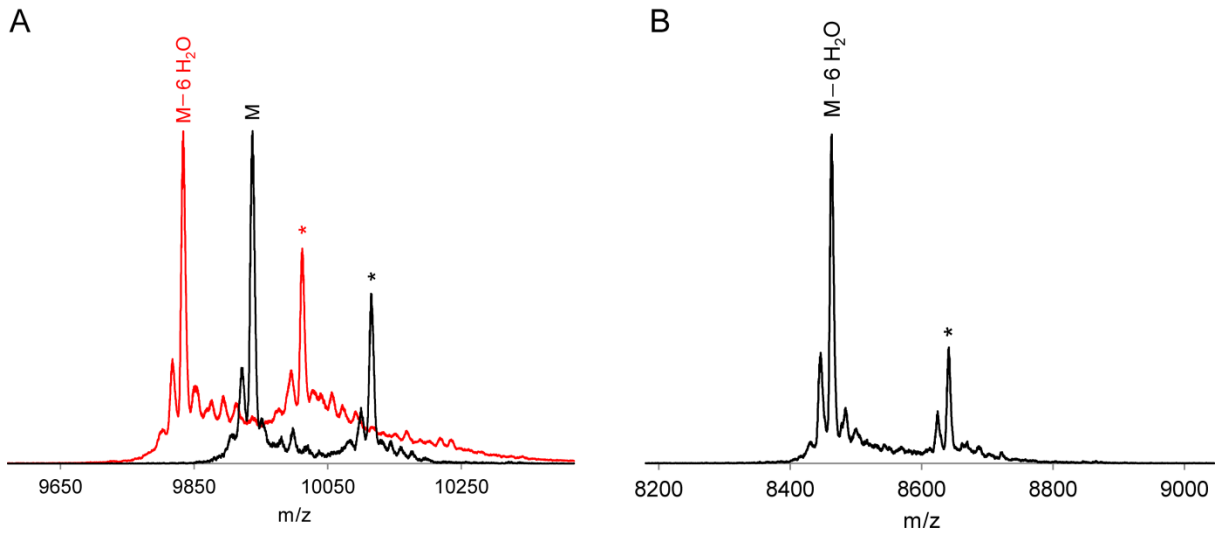

Figure 2.2. Modification of the FlvA peptides as observed by MALDI-TOF MS analysis. Modification of the FlvA peptides was observed as a loss of an integer number of 18 Da from the calculated, unmodified peptide mass. (A) In vitro modification of FlvA1.a by FlvM1. The red spectrum was observed upon exposure of FlvA1.a to FlvM1,

Figure 2.2. (cont.)

whereas the black spectrum was observed in the absence of FlvM1. FlvA1.a six-fold dehydrated, calculated average $[M+H]^+$ mass: 9832.74 m/z, observed average mass: 9833.65 m/z. FlvA1.a unmodified, calculated average $[M+H]^+$ mass: 9941.07 m/z, observed average mass: 9937.60 m/z. For FlvA1.a and FlvM1 heterologous co-expression, see Figure 2.3B. (B) In vivo modification of heterologously expressed FlvA2.b peptide (for the other FlvA2.x peptides, see Figure 2.4). FlvA2.b six-fold dehydrated, calculated average $[M+H]^+$ mass: 8463.51 m/z, observed mass: 8463.42 m/z. * Denotes ions with masses that correspond to gluconoylated starting material or product (57).

Co-expression of FlvA2.g with FlvM2 resulted in dehydrated peptides whereas co-expression of FlvA2.g with FlvM1 yielded no dehydration as shown by the corresponding MALDI-TOF MS spectra (Figure 2.3A). Further experiments with all of the FlvA2.x peptides showed they were also dehydrated upon co-expression with FlvM2 in *E. coli*. Similarly, FlvA1.a was dehydrated upon co-expression with FlvM1 (Figure 2.3B). Identical results were obtained in vitro with enzymes and substrates that were expressed as His_6-tagged fusion proteins and purified by immobilized affinity chromatography (IMAC, e.g. Figure 2.2A). In all cases the number of dehydrations was observed to be less than the total number of Ser and Thr residues within the core peptide (Table 2.1, Figure 2.2B, Figure 2.4), but Ser and Thr residues that escape dehydration are not unusual in lanthipeptides (58) including the lichenicidins (35, 36).

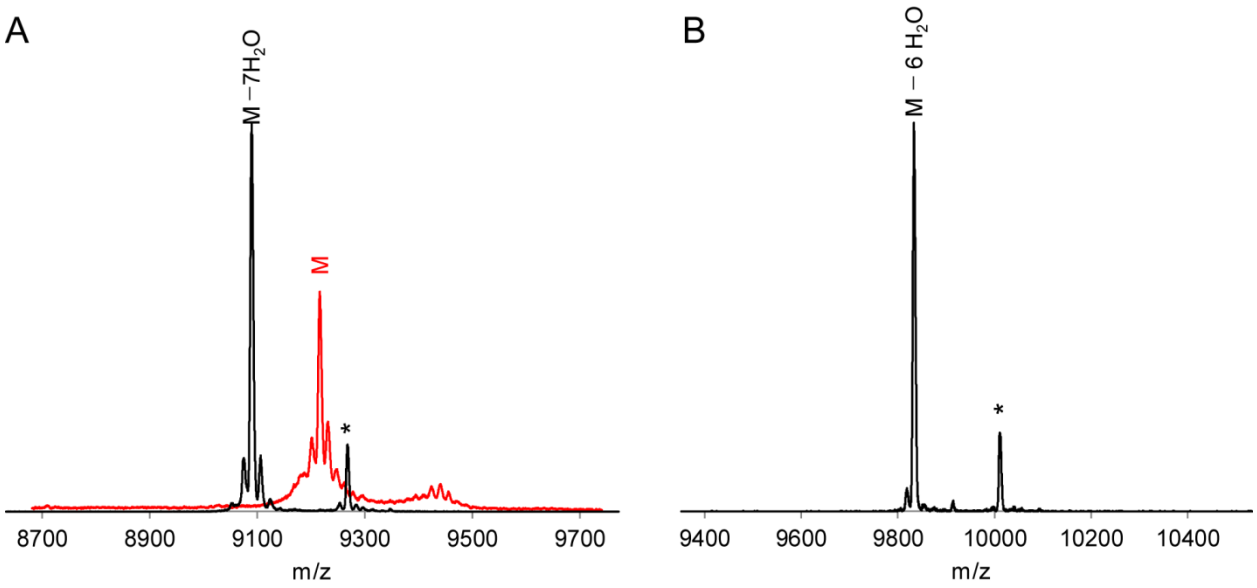

Figure 2.3. (A) Co-expression of FlvA2.g with FlvM1 and FlvM2. The observed major dehydration product is indicated above each peak. Co-expression of FlvA2.g with FlvM2, but not FlvM1 resulted in modified peptide. The red spectrum was observed upon co-expression of FlvA2.g with FlvM1; the black spectrum was observed upon co-expression of FlvA2.g with FlvM2. FlvA2.g unmodified, calculated (calc.) average (avg.) $[M+H]^+$ mass: 9213.4 m/z, observed (obsv.) avg. mass: 9214.6 m/z. FlvA2.g seven-fold dehydrated, calc. avg. $[M+H]^+$ mass: 9087.3, obsv. avg. mass: 9088.9 m/z. (B) MALDI-TOF-MS spectra of full length FlvA1.a after co-expression with FlvM1. FlvA1.a six-fold dehydrated, calc. avg. $[M+H]^+$ m/z: 9832.97, obsv. avg. m/z: 9832.63.

Table 2.1. Extent of dehydrations observed after co-expression of FlvA2.x with FlvM2 and FlvA1.a with FlvM1. The yields obtained for the core portion of the modified FlvA peptides range from 100-800 µg of purified peptide from 2 L of *E. coli* culture, depending on the method of removal of the leader peptide.

Substrate	Major product	Number of Ser and Thr in the core peptide	Number of Cys in the core peptide	Lan	MeLan
FlvA2.a	−7 H_2O	8	1	DL	-
FlvA2.b	−6 H_2O	9	3	-	DL
FlvA2.c	−5 H_2O	10	4	LL	DL
FlvA2.d	−7 H_2O	9	5	DL, LL	DL
FlvA2.e	−7 H_2O	10	4	LL	DL
FlvA2.f	−4 H_2O	5	2	-	DL
FlvA2.g	−7 H_2O	9	4	LL	DL
FlvA2.h	−7 H_2O	10	4	DL, LL	DL
FlvA1.a	−6 H_2O	8	6	DL	DL

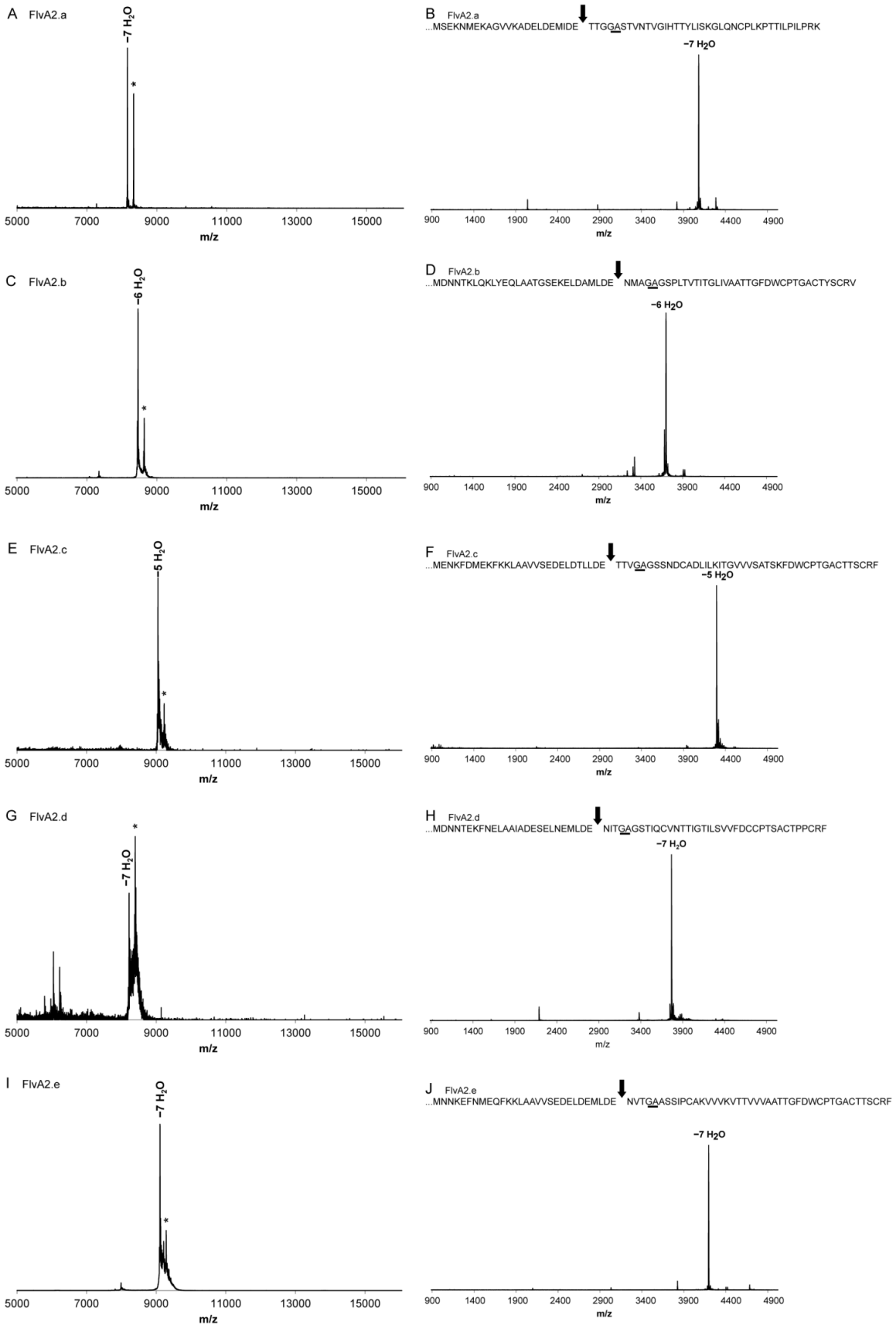

Figure 2.4. (cont.)

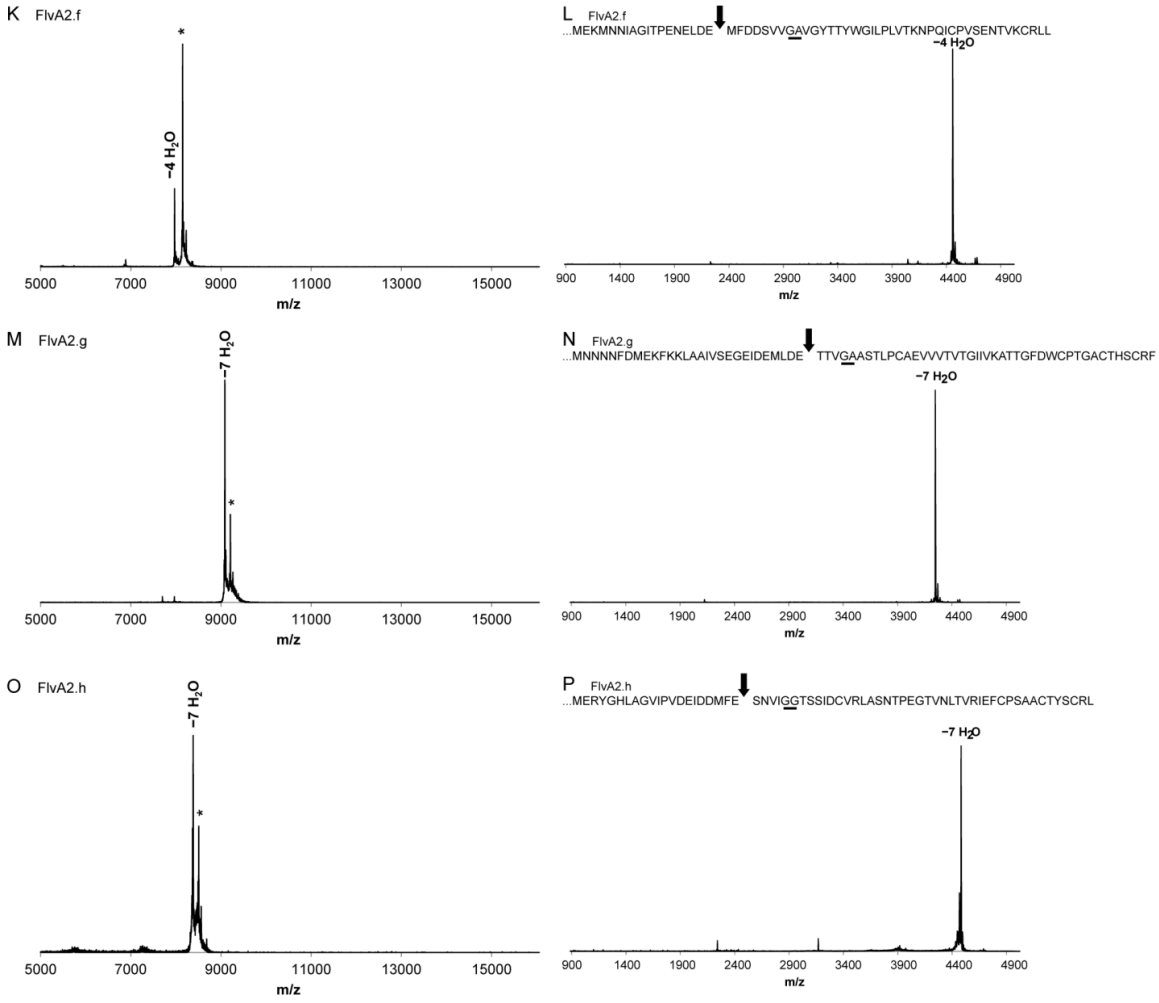

Figure 2.4. MALDI-TOF-MS spectra of full length FlvA2.x peptides following co-expression with FlvM2 (panels A, C, E, G, I, K, M, O). The corresponding FlvA2.x fragments formed upon Glu-C digestion and RP-HPLC purification are shown in panels B, D, F, H, J, L, N, P. The arrow indicates the position at which endoproteinase Glu-C cleaved. * Denotes ions with masses that correspond to gluconoylated starting material or product. The observed major dehydration product is indicated above each peak. (A, B) FlvA2.a full length, 7-fold dehydrated, calc. avg. $[M+H]^+$ m/z: 8156.33, obsv.: 8157.87. Glu-C fragment calc. $[M+H]^+$ avg. m/z: 4080.91, obsv.: 4080.58. (C, D) FlvA2.b 6-fold dehydrated full length, calc. avg. $[M+H]^+$ m/z: 8463.51 m/z, obsv.: 8463.42. Glu-C fragment calc. avg. $[M+H]^+$ m/z: 3700.34, obsv.: 3700.29. (E, F) FlvA2.c full length, 5-fold dehydrated, calc. avg. $[M+H]^+$ m/z: 9048.18, obsv.: 9049.10. Glu-C fragment calc. avg. $[M+H]^+$ m/z: 4295.96, obsv.: 4295.16 m/z. (G, H) FlvA2.d full length, 7-fold dehydrated, calc. avg. $[M+H]^+$ m/z: 8217.13, obsv.: 8217.97. Glu-C fragment calc. avg. $[M+H]^+$ m/z: 3767.45, obsv.: 3767.00. (I, J) FlvA2.e full length, 7-fold dehydrated, calc. avg. $[M+H]^+$: 9107.32, obsv.: 9108.08 m/z. Glu-C fragment calc. avg. $[M+H]^+$: 4195.97, obsv.: 4195.46 m/z. (K, L) FlvA2.f full length, 4-fold

Figure 2.4. (cont.)

Figure 2.4. (cont.)

dehydrated, calc. avg. $[M+H]^+$ m/z: 7970.05, obsv.: 7970.78 m/z. Glu-C fragment calc. avg. $[M+H]^+$ m/z: 4458.30, obsv.: 4458.79 m/z. (M, N) FlvA2.g full length, 7-fold dehydrated, calc. avg. $[M+H]^+$ m/z: 9087.29, obsv.: 9087.93 m/z. Glu-C fragment calc. avg. $[M+H]^+$ m/z: 4248.00, obsv.: 4248.33 m/z. (O, P) FlvA2.h full length, 7-fold dehydrated, calc. avg. $[M+H]^+$ m/z: 8382.40, obsv.: 8383.19 m/z. Glu-C fragment calc. avg. $[M+H]^+$ m/z: 4482.17, obsv.: 4482.77 m/z. The "…" preceding each Glu-C fragment corresponds to the amino acid sequence (GSSHHHHHHSQDP) arising from translation of plasmid RSFDuet-1 at the start codon of MCS1.

2.2.4 FlvM1 is selective for FlvA1 and FlvM2 is selective for the FlvA2 peptides

To assess the substrate tolerance of FlvM1, FlvA2.a and FlvA2.g were also incubated in vitro with FlvM1 under identical conditions. Analysis of the assay revealed little to no modification of either peptide by FlvM1 (Figure 2.5). Similarly, FlvM2 was unable to modify FlvA1.a to any appreciable extent when exposed to the peptide *in vitro* under the standard reaction conditions (Figure 2.5). The inability of FlvM1 to modify FlvA2.a and FlvA2.g and FlvM2 to modify FlvA1.a suggests a high degree of substrate selectivity that is probably governed by recognition of the different leader peptides (Figure 2.1B) (59, 60).

2.2.5 Partial FlvA2.x leader peptide removal and structural analysis

The alignment of the amino acid sequences of the FlvA peptides shows a conserved GA/GG motif located in the middle of the peptides (Figure 2.1B). Based on previous studies of other class II lanthipeptides, this motif likely marks the boundary between the FlvA leader and core peptides (61). Usually, leader peptides ending in the GA/GG motif are removed by the Cys protease domain of LanT transporters (62), and indeed such a protein is present in the gene cluster (FlvT, Figure 2.1A). Evidence that the GA/GG sequence is indeed the cleavage motif for the FlvA peptides was provided by co-expression of FlvT with FlvM2 and FlvA2.g in *E. coli*. The supernatant of the co-expression was analyzed by MALDI-TOF MS, which resulted in detection of masses corresponding to cleavage of modified FlvA2.g after the GA sequence (see Section 2.2.6). These masses were not observed in the culture co-expressing only FlvM2 and FlvA2.g. This observation is consistent with the previously reported removal of the leader peptide of modified LicA2 upon expression of the *lic* gene cluster, which contains *licT*, in *E. coli* (36). Although the use of FlvT confirmed the predicted leader peptide cleavage site, this strategy

unfortunately did not yield sufficient quantities of the Flvα and β-peptides for structure determination, in part because of the difficulty associated with purifying the desired, modified peptide from partially dehydrated peptides that were also secreted.

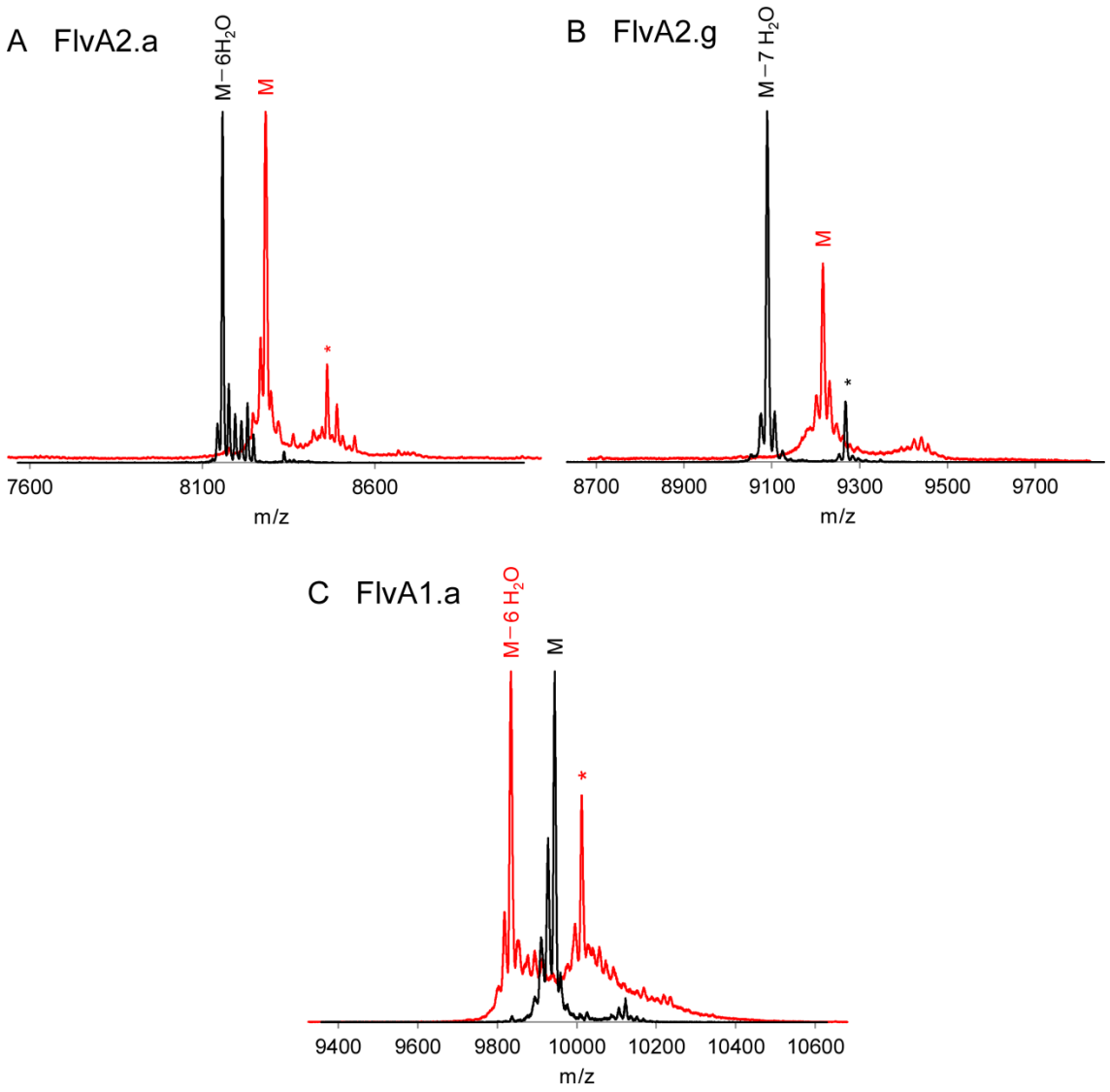

Figure 2.5. FlvM1 specifically modifies FlvA1 and FlvM2 modifies FlvA2. ***** Denotes ions with masses that correspond to gluconoylated starting material or product (57). (A) In vitro modification of FlvA2.a by FlvM2 (black), but not FlvM1 (red). FlvA2.a unmodified, calc. avg. $[M+H]^+$ m/z: 8282.4, obsv.: 8282.2. FlvA2.a 7-fold dehydrated, calc. avg. $[M+H]^+$ m/z: 8156.3, obsv.: 8157.9. (B) In vitro modification of FlvA2.g by FlvM2 (black), but not FlvM1 (red). FlvA2.g unmodified, calc. avg. $[M+H]^+$ m/z: 9213.4, obsv.: 9214.6. FlvA2.g 7-fold dehydrated, calc. avg. $[M+H]^+$ m/z: 9087.3, obsv.: 9088.9. (C) In vitro modification of FlvA1.a by FlvM1 (red), but not FlvM2

Figure 2.5. (cont.)

(black). FlvA1.a unmodified, calc. avg. $[M+H]^+$ m/z: 9941.1, obsv.: 9944.2. FlvA1.a 6-fold dehydrated, calc. avg. $[M+H]^+$ m/z: 9833.0, obsv.: 9833.7.

We therefore turned towards in vitro removal of the putative leader peptides from the His_6-tagged FlvA2.x peptides after co-expression with FlvM2 and purification by IMAC. The purified peptides were first subjected to proteolysis by endoproteinase Glu-C. Although FlvA2.f-2.h have Glu residues within their core peptides, appreciable proteolysis at these sites was not observed, probably because they are protected by the post-translational modifications. Instead, treatment with Glu-C resulted primarily in fragments arising from proteolysis within the leader peptide sequence, specifically at Glu−5 for FlvA2.a, Glu−6 for FlvA2.b through FlvA2.e and FlvA2.g, Glu−7 for FlvA2.h, and Glu−10 for FlvA2.f (for sequences, see Figure 2.1B). The Glu-C-generated C-terminal fragments of the modified FlvA2.x peptides were purified by reversed-phase high performance liquid chromatography (RP-HPLC) and analyzed by electrospray ionization tandem mass spectrometry (ESI-MS/MS) (e.g. Figure 2.6A).

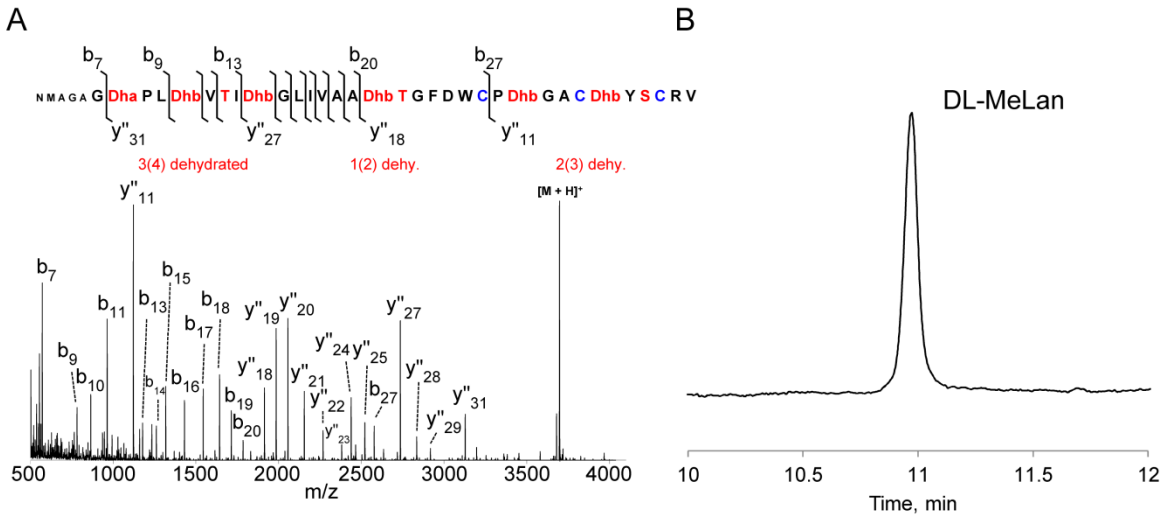

Figure 2.6. Structural analysis of the FlvM2-modified FlvA2.b. (A) Tandem mass spectrum of the modified FlvA2.b peptide after treatment with the endoprotease Glu-C. $[M+H]^+$ denotes the peak corresponding to the protonated, monoisotopic parent mass. Fragment ions were not observed at the C-terminus of the peptide. For tandem mass spectra of the remaining FlvA peptides, see Figure 2.7. (B) GC/MS, selected ion monitoring at 379 Da, indicating the presence of MeLan. The identity of the eluting material was confirmed by co-elution with an authentic MeLan standard (Figure 2.8B).

A lack of fragmentation was observed in the C-termini of all of the modified FlvA2.x peptides analyzed, suggesting the presence of thioether rings (Figure 2.6A and 2.7). ESI-MS/MS analysis of modified FlvA1.a also indicated the absence of fragmentation in the C-terminus of the peptide, consistent with cyclization in this region of the peptide (Figure 2.7). To confirm the presence of (Me)Lan residues within the FlvM-modified FlvA1.a and FlvA2.x peptides, the peptides were hydrolyzed and the resulting residues were derivatized to the corresponding pentafluoropropionamide methyl esters. These volatile derivatives were then separated by chiral gas chromatography monitored by mass spectrometry (GC/MS), which confirmed that the modified FlvA1.a and FlvA2.x peptides indeed contained (Me)Lan residues (Figure 2.6B and Figure 2.8). The stereochemistry of the observed (Me)Lan residues was determined by using synthetic standards with known stereochemistries (63), and is listed in Table 2.1. As anticipated based on previous predictions (43), peptides containing a Dhx-Dhx-Xxx-Xxx-Cys motif resulted in LL-Lan whereas all other Lan and MeLan residues had the DL stereochemistry.

The possibility of incomplete cyclization was assessed by iodoacetamide (IAA) assays; iodoacetamide selectively alkylates free Cys residues and not thioethers. For all FlvM-modified peptides except FlvA1.a, treatment with IAA did not result in appreciable formation of adducts, which is indicative of complete cyclization; di-alkylation was observed with FlvM1-modified FlvA1.a (Figure 2.9C). Based on the fragmentation pattern (Figure 2.9D) and the alignment with other α-peptides (Figure 2.1C), the di-alkylation product most likely involves alkylation at Cys8 and Cys29. Two non-cyclized Cys are also present in haloduracin α, but the Cys residues involved do not align with those of Flvα.a. The two free Cys residues in modified FlvA1.a presented the possibility of a disulfide bond in the structure of mature FlvAα.a. The antimicrobial activity of some lantibiotics has been reported to depend on the presence of such disulfide bonds (53, 64-66). Modified FlvA1.a was therefore incubated with oxidized and reduced glutathione in an attempt to oxidatively fold the peptide (67). However, the peptide remained reduced under these conditions, as observed by ESI-MS, suggesting that unlike lantibiotics that contain disulfide bonds, the two Cys in modified FlvA1.a are not in a conformation in which they readily form a disulfide bond (51, 54). We cannot rule out that disulfide bond formation needs to occur prior to (Me)Lan installation.

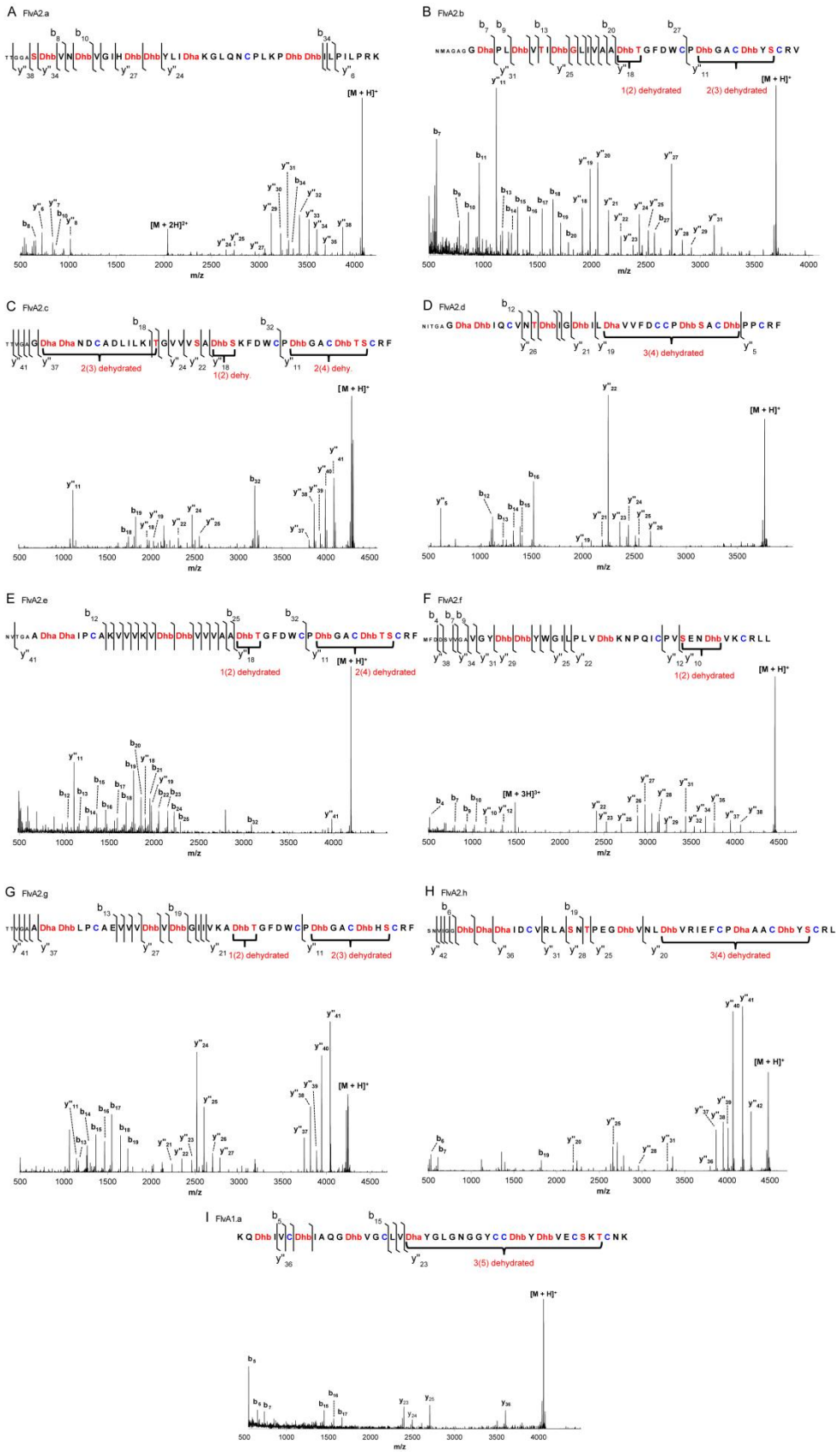

Figure 2.7. (cont.)

Figure 2.7. Tandem MS spectra of the FlvM modified FlvA peptides. $[M+H]^+$ denotes the protonated, monoisotopic mass of the parent peak. The dehydration pattern supported by the tandem MS data is indicated. For regions in which the dehydration pattern is ambiguous, the number of dehydrations supported by the data is indicated with the total possible number of dehydrations in parentheses. For example, "1(2) dehydrated" means one out of two of the residues for the particular region of peptide were dehydrated. (A) FlvA2.a, (B) FlvA2.b, (C) FlvA2.c, (D) FlvA2.d, (E) FlvA2.e, (F) FlvA2.f, (G) FlvA2.g, (H) FlvA2.h, and (I) FlvA1.a.

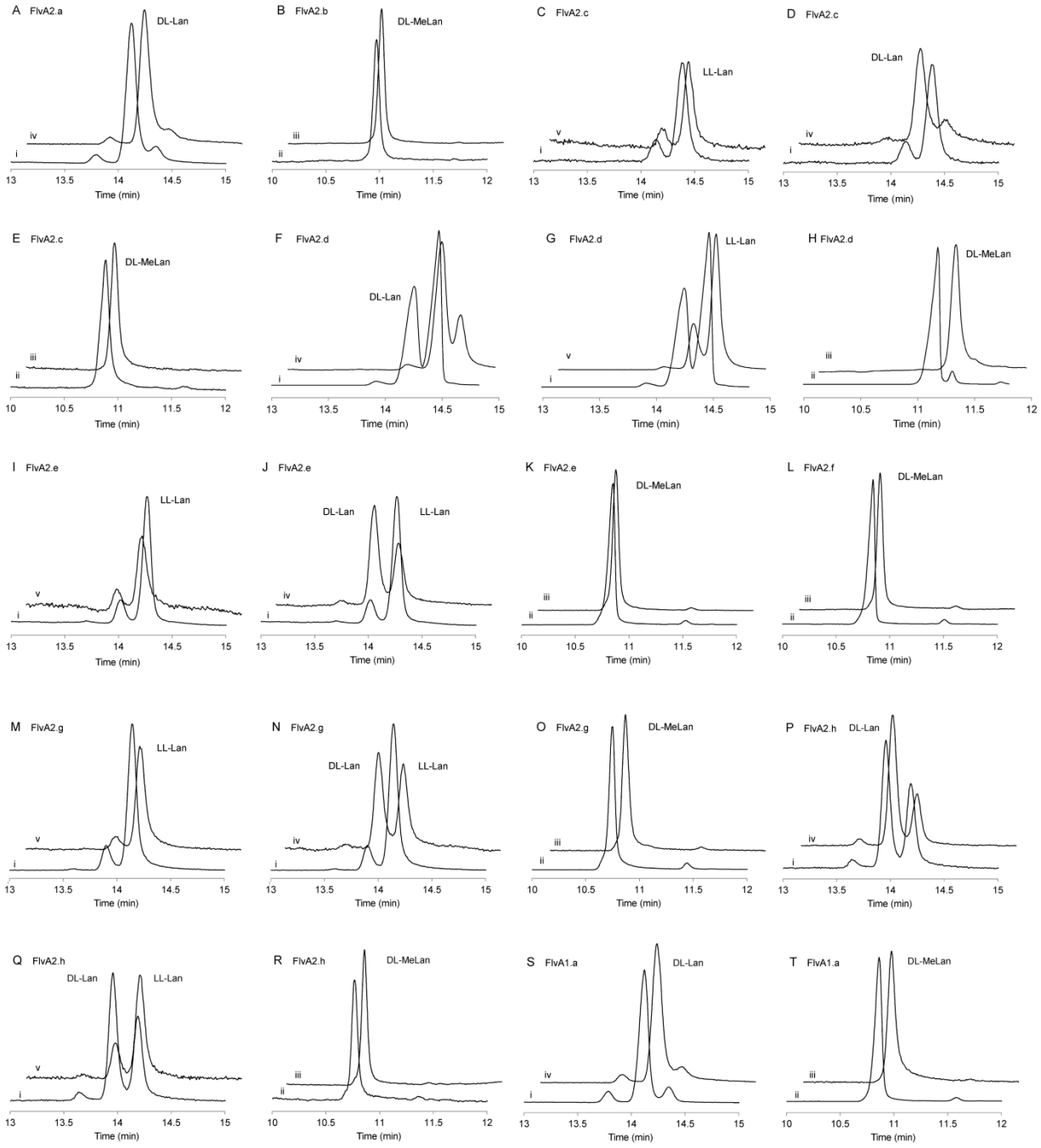

Figure 2.8. GC/MS, selected ion monitoring (SIM) at 365 Da and 379 Da of the FlvM modified FlvA peptide fragments obtained by treatment with Glu-C, indicating the presence and stereochemistry of Me(Lan) residues. The stereochemistry of the (Me)Lan residues was confirmed by co-elution with authentic standards, (A) FlvA2.a. (B) FlvA2.b. (C-E) FlvA2.c. (F-H) FlvA2.d. (I-K) FlvA2.e. (L) FlvA2.f. (M-O) FlvA2.g. (P-R) FlvA2.h. (S, T) FlvA1.a. Foreground plots are of the sample while background plots are of sample mixed with standard. (i) SIM at 365 Da of the sample (Lan), (ii) SIM at 379 Da of the sample (MeLan), (iii) SIM at 379 Da of the sample mixed with DL-

Figure 2.8. (cont.)
MeLan standard, (iv) SIM at 365 Da of the sample mixed with DL-Lan standard, (v) SIM at 365 Da of the sample mixed with LL-Lan standard. When no data is shown for Lan or MeLan, it indicates that the sample did not contain the residue.

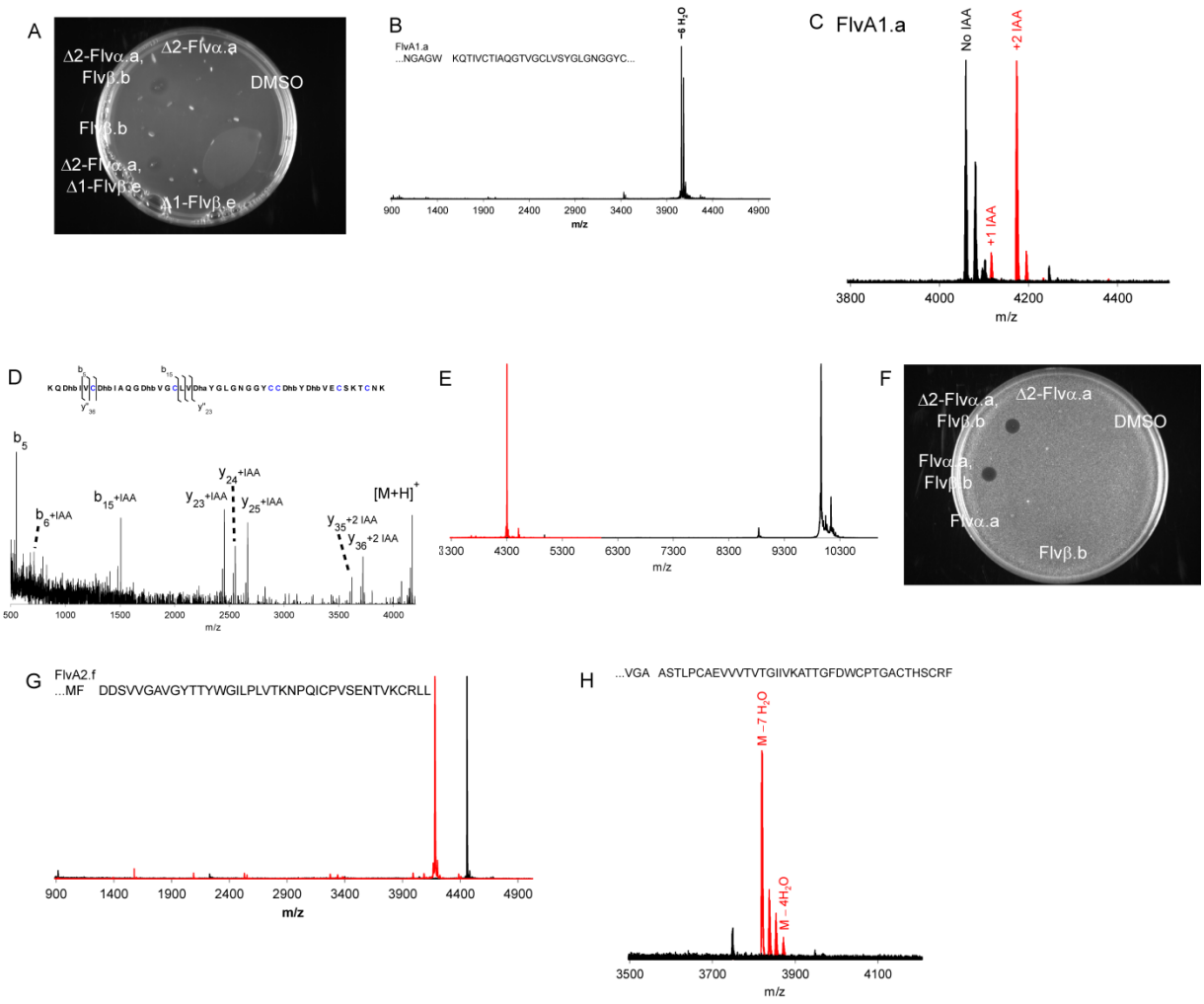

Figure 2.9. (A) Bioactivity of FlvA peptides on *Ruminococcus albus* 7. Very similar activity was also observed with *Ruminococcus flavefaciens* C94. (B) Chymotrypsin treatment of modified FlvA1.a. The observed major peak corresponded to six-fold dehydrated FlvA1.a, calc. avg. $[M+H]^+$ m/z: 4060.80, obsv.: 4060.68. The "…" before and after the amino acid sequence represents the omitted portion of the leader and core peptides, respectively GSSHHHHHHSQDPMNKNPIYRSEEEAKDIACGNVAAELDENSQALDAI and CTYTVE-CSKTCNK. (C) Iodoacetamide (IAA) treatment of FlvA1.a. The number of alkylation adducts is indicated above each peak, resulting in an increase of 57 Da. The red spectrum was observed upon exposure to iodoacetamide, whereas the black spectrum was observed in the absence of iodoacetamide. FlvA1.a six-fold dehydrated calc. avg. $[M+H]^+$ m/z: 4060.80, obsv.: 4059.02, obsv. mono-alkylated avg. mass: 4115.81 m/z, and obsv. di-alkylated avg. mass: 4172.75 m/z. (D) Tandem MS spectrum of alkylated FlvA1.a. $[M+H]^+$ denotes the protonated, monoisotopic mass of the parent peak. The alkylation pattern supported by the tandem MS data is indicated. (E) Modification and proteolysis of FlvA1.a. The black spectrum was observed upon co-expression of FlvA1.a(A−1insE) and FlvM1. The red spectrum was observed upon exposure of the modified FlvA1.a(A−1insE) peptide to Glu-C. FlvA1.a(A−1insE) six-fold dehydrated calc. avg. $[M+H]^+$ m/z: 9962.09, obsv.: 9962.42. FlvAα.a calc. avg. $[M+H]^+$ m/z: 4304.07, obsv.:

Figure 2.9. (cont.)

4303.29. (F) Bioactivity of Δ2-Flvα.a, Flvα.a, and Flvβ.b. (G) Aminopeptidase treatment of modified FlvA2.f resulting in +7Flvβ.f. The red spectrum was observed upon treatment of the modified FlvA2.f Glu-C fragment with aminopeptidase, whereas the black spectrum was observed in the absence of aminopeptidase. The observed major peak corresponds to 4-fold dehydrated FlvA2.f, calc. avg. $[M+H]^+$ m/z: 4179.92, obsv.: 4180.98. The "…" preceding the amino acid sequence represents the omitted portion of the leader peptide that was removed using Glu-C GSSHHHHHHSQDPMEKMNNIAGITPENELDEMF. (H) Co-expression of FlvA2.g, FlvM2, and FlvT. The red spectrum was observed from the supernatant of *E. coli* cells in which FlvA2.g, FlvM2, and FlvT were co-expressed, whereas the black spectrum was observed from the supernatant of the FlvA2.g and FlvM2 co-expression. The observed major peak corresponded to seven-fold dehydrated FlvA2.g, calc. avg. $[M+H]^+$ mass: 3818.53 m/z, obsv. avg. mass: 3819.04 m/z. The "…" represents the omitted portion of the leader peptide GSSHHHHHHSQDP MNNNNFDMEKFKKLAAIVSEGEIDEMLDETT.

2.2.6 FlvT proteolyzes and exports the FlvA2.x peptides and dehydrated LicA2

Although insufficient quantities of isolable core peptide resulted from co-expression of FlvT with the FlvM2 and FlvA2.g (Figure 2.9H), the ability of FlvT to process the other FlvA2.x peptides remained unclear. Thus, the remaining FlvM2 and A2.x pairs were co-expressed with FlvT and in almost all cases, partially and fully dehydrated Flv2.x peptides were detected in the supernatant of the co-expression cultures (Figure 2.10). Notably, no fully modified peptide was observed in the co-expression culture of FlvA2.a. Additionally, co-expression of FlvT with LicA2 and LicM2 resulted in detection of dehydrated LicA2 in the supernatant of the expressing *E. coli* (Figure 2.10H). In contrast, FlvT co-expressed with FlvM1 and FlvA1.a did not result in detectable amounts of modified FlvA1.a. However, attempts to reconstitute the in vitro proteolytic activity of the Cys protease domain of FlvT (His_6-FlvT151) were unsuccessful despite soluble expression of the truncated protein.

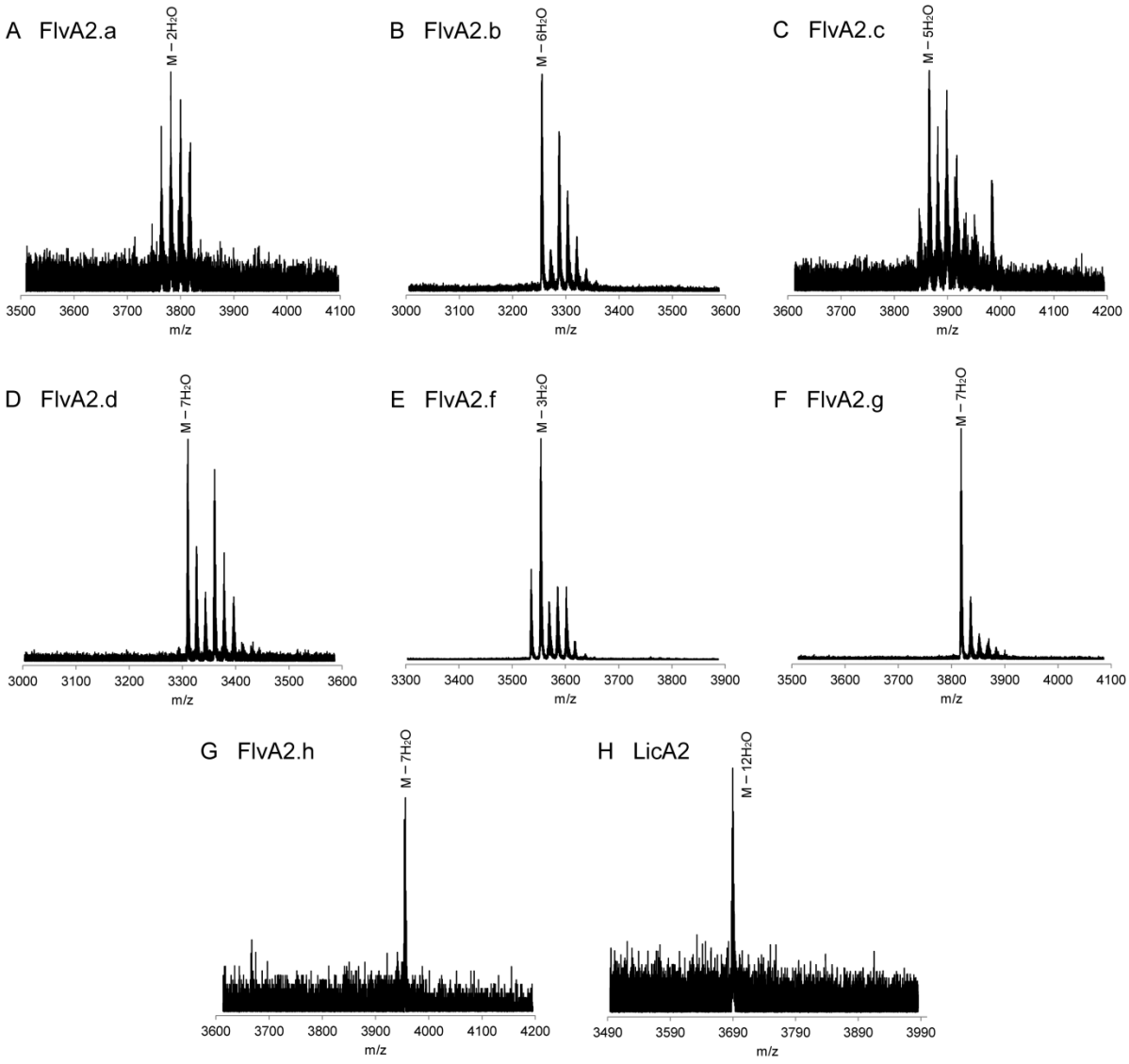

Figure 2.10. Modified core peptides observed in the supernatants of FlvT co-expression cultures. The masses of the labeled peaks corresponded to (A) two-fold dehydrated FlvA2.a, calc. avg. $[M+H]^+$ m/z: 3783.6, obsv.: 3781.8, (B) six-fold dehydrated FlvA2.b, calc. avg. $[M+H]^+$ m/z: 3255.8, obsv.: 3255.4, (C) five-fold dehydrated FlvA2.c, calc. avg. $[M+H]^+$ m/z: 3866.5, obsv.: 3865.6, (D) seven-fold dehydrated FlvA2.d, calc. avg. $[M+H]^+$ m/z: 3310.4, obsv.: 3311.0, (E) three-fold dehydrated FlvA2.f, calc. avg. $[M+H]^+$ m/z: 3554.3, obsv.: 3553.9, (F) seven-fold dehydrated FlvA2.g, calc. avg. $[M+H]^+$ m/z: 3818.5, obsv.: 3818.1, (G) seven-fold dehydrated FlvA2.h, calc. avg. $[M+H]^+$ m/z: 3954.6, obsv.: 3956.0, and (H) 12-fold dehydrated LicA2, calc. avg. $[M+H]^+$ m/z: 3690.2, obsv.: 3689.7.

2.2.7 Proteolytic removal of the remaining portion of the FlvA2.x leader peptides

Complete removal of the leader peptide was challenging as introduction of artificial protease cleavage sites resulted in either incomplete modification by the FlvM enzymes, incomplete proteolysis because of the post-translational modifications near the cleavage site, or cleavage in the core peptides, which are problems that have been previously observed for other lanthipeptides (55, 68, 69). We therefore resorted to treating the Glu-C-generated FlvA2.x peptide fragments with aminopeptidase (70-72). It was anticipated that the proteolytic activity of aminopeptidase would terminate upon encountering the dehydrated residues featured at or near the predicted N-termini of most of the FlvA2.x peptides (Figure 2.12). MALDI-TOF-MS analysis of aminopeptidase reactions indicated that for FlvA2.b, the predicted native product was formed. For FlvA2.c, FlvA2.e, and FlvA2.g, aminopeptidase removed all remaining residues of the leader peptides as well as the predicted first residue of the core peptides, resulting in N-terminal lanthionines. We term these products Δ1-Flvβ.c, Δ1-Flvβ.e, and Δ1-Flvβ.g. For FlvA2.d and FlvA2.h, the aminopeptidase treatment left two amino acids of the leader peptide on the final products, likely because of the high Gly content at the junction between the core and leader peptides, making these peptides less efficient substrates for aminopeptidase (73). We term these products +2Flvβ.d and +2Flvβ.h. For FlvA2.a, three residues remained after aminopeptidase treatment, resulting in +3Flvβ.a. Finally, in the case of FlvA2.f, seven residues from the leader peptide remained after aminopeptidase treatment (Figure 2.9G; +7Flvβ.f).

This sequential digestion approach works reasonably well when the N-terminus of the lanthipeptide is blocked by post-translational modifications, but if the N-terminus is a linear sequence as predicted for the product of FlvA1.a, the method does not work. Instead, removal of the FlvA1.a leader peptide was achieved via a single step proteolysis using chymotrypsin (Figure 2.9B). Assuming that the GA/GG sequence marks the end of the leader peptide, chymotrypsin cleaves after Trp2 in the core peptide of FlvM1-modified FlvA1.a, and therefore the resulting peptide was designated Δ2-Flvα.a. In an effort to avoid removal of the two N-terminal amino acid residues of the FlvA1.a core peptide, a mutant FlvA1.a peptide with Glu inserted before the first residue of the core peptide was generated (termed FlvA1.a(A−1insE)), and co-expressed with FlvM1. Doing so resulted in six-fold dehydrated FlvA1.a(A−1insE), and treatment of this mutant peptide with Glu-C yielded the desired Flvα.a (Figure 2.9E).

2.2.8 Antimicrobial activity of the modified FlvA core peptides

The peptides produced by in vitro leader peptide removal were purified by RP-HPLC, and the modified FlvA peptides were first assayed for antimicrobial activity against the lantibiotic-sensitive, aerobic microorganism *Micrococcus luteus* DSM 1790. In an agar diffusion test, Δ1-Flvβ.c and Δ1-Flvβ.e exhibited antimicrobial activity that was not synergistically enhanced by Δ2–Flvα.a (Figure 2.11). In contrast, combinations of Flvβ.b or Δ1-Flvβ.g with Δ2–Flvα.a displayed the synergistic antimicrobial activity that is characteristic of two-component lantibiotics. A comparable level of synergistic, antimicrobial activity was also observed for a combination of Flvα.a and Flvβ.b (Figure 2.9F), suggesting that the removal of the additional two amino acids at the N-terminus of Δ2–Flvα.a is not detrimental for bioactivity. Since we could obtain considerably larger quantities of Δ2–Flvα.a, this peptide was used for all subsequent bioassays. Δ1-Flvβ.g also displayed activity by itself that was lower than that observed when spotted with Δ2–Flvα.a. The remainder of the Flvβ-peptides (+3Flvβ.a, +2Flvβ.d, +7Flvβ.f, and +2Flvβ.h) did not display antimicrobial activity, whether tested separately or in combination with Δ2–Flvα.a (Figure 2.11). Enhanced activity beyond what was observed in pairwise combination was also not detected by combining all peptides. In light of their antimicrobial activity, Flvα.a, β.b, β.c, β.e, and β.g were designated flavecins.

M. luteus is often used as an indicator strain to test for antimicrobial activity of lipid II-targeting peptides such as two-component lantibiotics, but it is not a rumen bacterium. To investigate the activity of the peptides against bacteria that would be more relevant in the ruminal context, the flavecins were assayed against *Ruminococcus albus* 7 and *R. flavefaciens* C94 under anaerobic conditions. Interestingly, very similar patterns of activity were observed with Δ2–Flvα.a, Flvβ.b, and Δ1-Flvβ.e (Figure 2.9A). As with *M. luteus*, the activity was weak, suggesting that these organisms are not the intended target of these compounds. Alternatively, the weak antimicrobial activity exhibited by the flavecins may suggest that these peptides mediate more subtle microbial interactions (74).

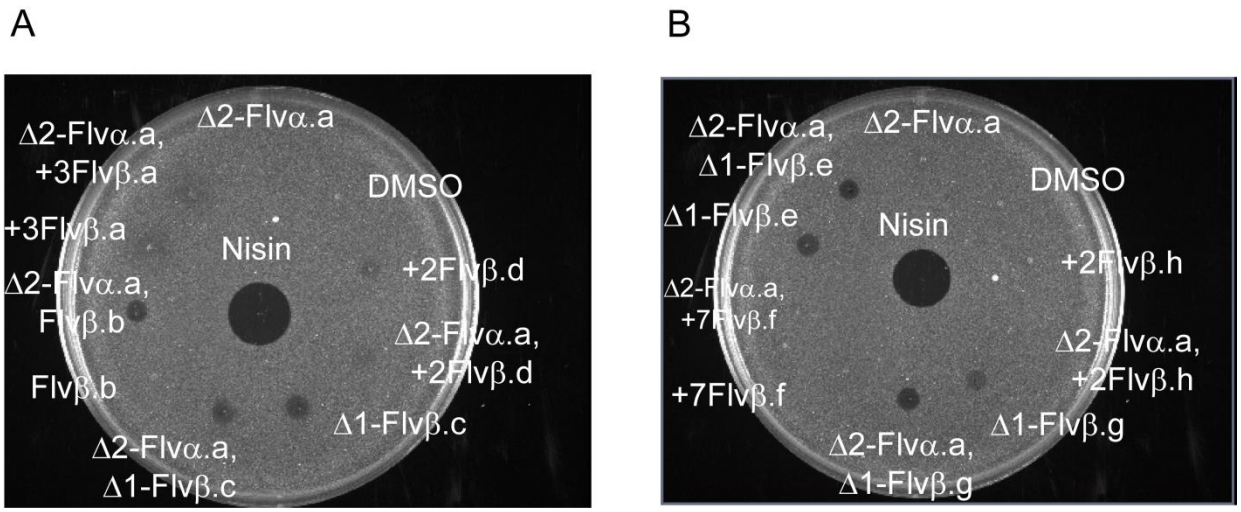

Figure 2.11. Bioassay of FlvM-modified FlvA core peptides against *M. luteus* DSM 1790. The modified Δ2-Flvα.a peptide was assayed at a concentration of 1 mM and the Flvβ.x peptides were assayed at a concentration of 0.5 mM. A volume of 1 µL was used for spots consisting of a single peptide while 0.5 µL of each solution was used for spots combining two peptides; nisin was assayed using 0.2 µL of a 100 µM solution.

2.3 Discussion

The sequence similarity of LanM enzymes was used to survey genomes for unusual class II lanthipeptide biosynthetic systems, which resulted in the identification of the novel *flv* gene cluster in *R. flavefaciens* FD-1. Recently, Singh and Sareen also bioinformatically identified the *flv* cluster via the sequence similarity of FlvT to HalT (75), but this previous study did not characterize the enzymes or the products of the cluster nor discussed its unusual characteristics. Some features of the *flv* cluster are reminiscent of two-component lantibiotics but the high number of substrate genes with diverse sequences is not consistent with a typical two-component system. Previous reports on antimicrobial peptides produced by species from the *Ruminococcus* genus (13, 17, 76) suggested that the mature FlvA peptides could represent antimicrobial defenses of *R. flavefaciens* FD-1. In addition, previous studies of the lantibiotics ruminococcin A and butyrivibriocin OR79A (12, 14, 77) suggested that anaerobic bacteria associated with the digestive system are a source of bioactive lanthipeptides. Thus, characterization of the *flv* gene cluster and the resulting modified FlvA peptides was undertaken.

Unlike most previously characterized lanthipeptides, Flv peptide production from the native organism was not detected in appreciable amounts. This observation could be attributed to a lack of *flv* gene cluster expression under the laboratory culture conditions. Some evidence for this hypothesis is presented in the case of the lantibiotic ruminococcin A for which transcription of the biosynthetic genes in the native organism *Ruminococcus gnavus* E1 is dependent upon growth phase and the presence of trypsin (78). The variety of genetic contexts and regulatory elements associated with lanthipeptide biosynthetic gene clusters, discussed in Chapter 1, entails optimization of culture conditions or genetic tools for each native organism in order to access the products of the cluster. In order to circumvent the lack of observed production of the Flv peptides from *R. flavefaciens* FD-1, the established heterologous host *E. coli* was used for Flv peptide production.

The structure of Flvα.a suggested by sequence homology and tandem MS data is consistent with those of other α-peptides of two-component systems. Although a general lack of fragmentation from the middle of the peptide to the C-terminus makes definitive assignment of the ring structure in this region difficult (Figure 2.7I), the ring pattern is very likely the same as those in Halα, Ltnα, and Licα (Figure 2.1C) (35, 42, 79). The most N-terminal Cys residue does not appear to be cyclized since fragmentation is observed on either side of the residue, and IAA assays showed the presence of two free Cys (Figure 2.7I and 2.9C). The possibility that these two Cys might be engaged in a disulfide bond was not supported by attempts to oxidatively cyclize them.

Based on the results obtained from tandem MS and GC/MS, the modified FlvA2 peptides can be divided into two groups (Figure 2.12). The first group features an N-terminal ring, and the second category does not have the residues required to form this N-terminal ring. The assignment of Flvβ-peptides into the former group was based on the lack of fragmentation at the N-terminal extremity of these peptides as well as the LL stereochemistry of the Lan rings observed by GC/MS. Such a ring is also present in the β-peptides of haloduracin and lichenicidin, although the stereochemistry of the latter has not been reported (35, 42, 79). The central portion of all Flvβ-peptides contains a ring consisting of seven residues. The assignment of this ring is based on the tandem MS data that show a clear lack of fragmentation for all Flvβ-peptides in this region (Figure 2.7). This ring also nearly aligns to the B ring of other two-component peptides, but the latter feature a smaller five-amino-acid ring in this region (Figure

2.1D). The size of this central ring is not entirely clear for Flvβ.b, β.c, β.e, and β.g since a six-amino-acid ring is also possible. However, formation of a seven-amino-acid ring is more likely based on Flvβ.a, β.d, β.f and β.h, which unambiguously contain such a ring. Also, comparison of the fragmentation patterns of Flvβ.b, β.c, β.e and β.g imply that one Ser/Thr within this region consistently escapes modification (Figure 2.12), suggesting the Ser/Thr that could result in a six-amino-acid ring (underlined in Figure 2.12) is actually not dehydrated.

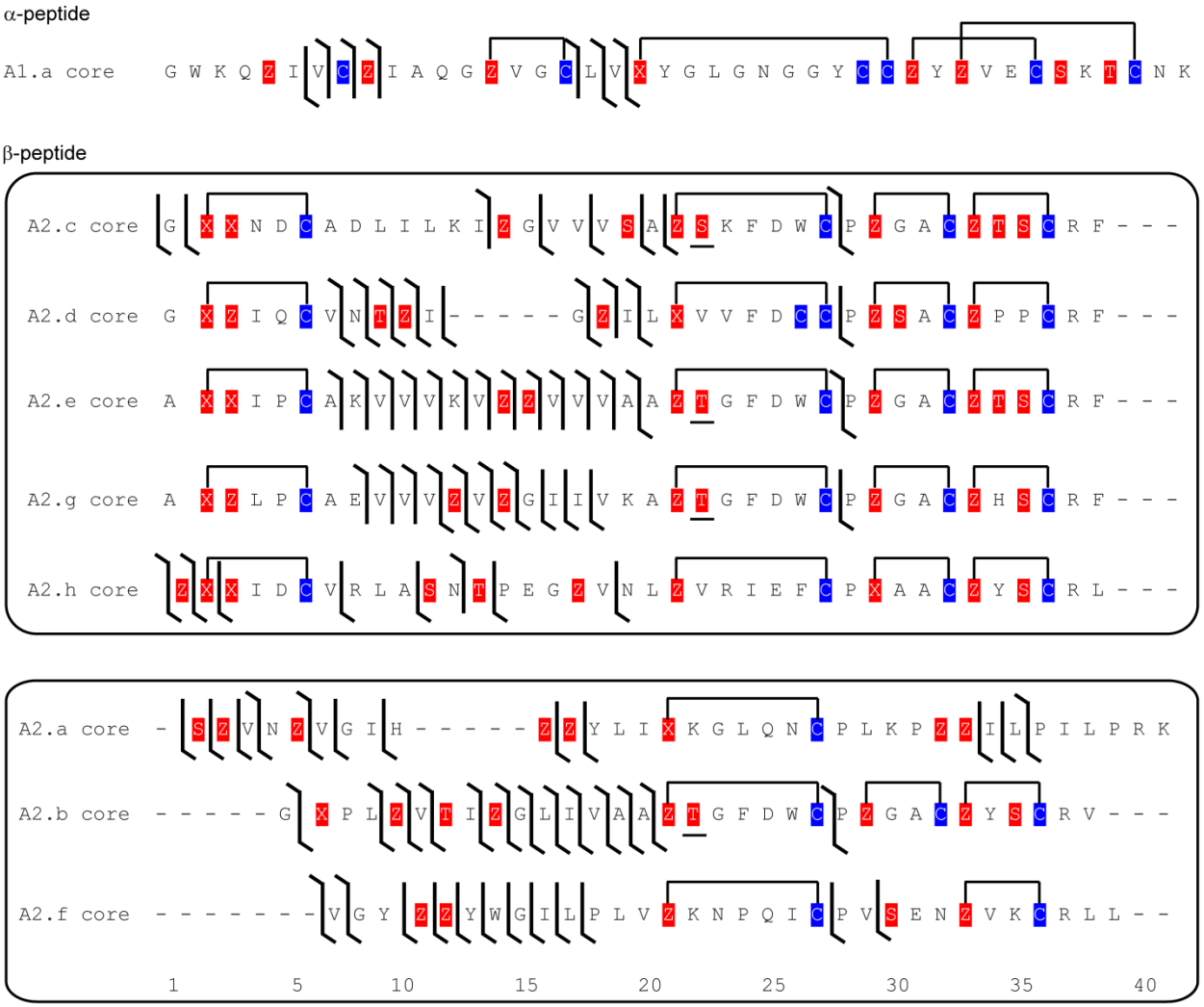

Figure 2.12. Tandem MS fragmentation patterns and proposed structures of the modified FlvA core peptides. The FlvA peptides are divided into the categories that are discussed in the text, and the alignments highlight the similar ring topology of the peptides in each category. X = Dha or Ala, when engaged in a Lan, and Z = Dhb or Abu, when engaged in a MeLan.

Despite the different ring topologies observed in the FlvA2.x peptides, the heterologously expressed FlvT protein was able to proteolytically remove the leader peptide and export most of the FlvA2.x peptides. Furthermore, varying amounts of dehydration of each FlvA2.x peptide were observed in the supernatant of the FlvT, FlvM2, and FlvA2.x co-expressions (Figure 2.10), suggesting appreciable tolerance by FlvT for secondary structural variations in the core peptide. The sequence tolerance exhibited by FlvT was further illustrated by its ability to export the dehydrated core peptide of the related lantibiotic lichenicidin β (Figure 2.10). Lastly, the lack of observed in vitro activity for the Cys protease domain of FlvT could possibly be remediated by optimizing the site of truncation. In vitro reconstitution of FlvT activity could present a broad substrate protease capable of processing two-component β peptide-like sequences. The in vitro reconstitution and subsequent investigations of LanT enzymes may present an enabling alternative to the current array of leader peptide removal strategies.

Two combinations of peptides resulted in synergistic bioactivity. Adding Δ2-Flvα.a to either Flvβ.b or Δ1-Flvβ.g resulted in clearly enhanced zones of growth inhibition. A comparison of the structures determined for Flvβ.b and Δ1-Flvβ.g indicates that the similarities are all in the C-terminal region, suggesting that this segment of the peptide is responsible for the interactions of the Flvβ peptides with the Flvα peptide. This observation is in agreement with the extensive mutagenesis data on the β-peptide of lacticin 3147 (80). Two possibilities were considered for the relatively low observed antibacterial activity. Perhaps the small deviations between the predicted leader peptide removal sites and the products obtained after in vitro leader peptide removal resulted in imperfect interactions of the peptides. We did not find this explanation very likely since the N-termini of the eight Flvβ-peptides (and other two-component lanthipeptides) are already highly heterogeneous by sequence (Figure 2.1D), and more importantly, the antimicrobial activity observed for the predicted Flvα.a and β.b core peptides was comparable to the combination of Δ2−Flvα.a and β.b (Figure 2.9F). Alternatively, perhaps the *flv* cluster is a pseudo cluster. We find this possibility also highly unlikely. It seems no coincidence that the four copies of the α-peptide are highly conserved, consistent with the general model of two-component lantibiotics in which this peptide recognizes a structurally conserved target (lipid II, see Chapter 1). The observation that the cluster encodes four such copies and eight copies of the β-peptides in light of a 1:2:2 stoichiometry observed for lipid II:Halα:Halβ (32) also seems non-coincidental, but attempts to obtain increased activity by various combinations of peptides have

thus far not been successful. The lack of robust antimicrobial activity seems especially puzzling given the presence of hallmark lipid II binding features in the Flvα.a peptide introduced in Chapter 1. However, since the activity of the two-component lantibiotics is dependent on the interaction of two separate polypeptides, the effects of differences in the primary structure of the Flvβ.x peptides, as compared to the β peptides of characterized two-component lantibiotics (Figure 2.1D), on antimicrobial activity are unclear.

The weak observed activity may instead reflect our inability to identify the physiological target organism(s). Ruminal two-component lantibiotics may be much more fine-tuned for specific target organisms than observed in other environments. It has been reported that bacterial diversity in vertebrate-associated communities is quite different from that in soil or aquatic environments, with 16S rRNA gene-based trees indicating extensive diversity at the species level rather than at higher levels as observed in the latter environments (81-84). Perhaps such an environment has promoted the thus-far unique example of a multi-component lantibiotic system with eight diverse β-peptides. The corollary of this hypothesis is that it may be very difficult to identify the specific physiological target organism of these peptides, especially if the flavecins exhibit the narrow target specificity that is not unusual for RiPPs (85, 86). In turn, this may explain the rather weak activity observed against the organisms tested thus far. At present, still very few lanthipeptide gene clusters are known from ruminant environments (87) and we have not observed any other two-component systems in the available genomes. As discussed below, this may be the result of the still limited available genome sequence information, and as more genomes are sequenced, additional such examples may be uncovered.

The current work on the genetic capacity of *R. flavefaciens* FD-1 to produce a diverse set of eight different β-peptides has similarities but also important differences compared to previous examples of combinatorial biosynthesis of lanthipeptides. In 2010, the first example was reported of a single enzyme making 30 different lanthipeptides in the cyanobacterium *Prochloroccocus* MIT 9313 (88). At the time this was a curious finding isolated to one genome, but in the intervening time as more genomes have been sequenced, this combinatorial biosynthesis has been found in many different cyanobacteria genera (25). To date, none of the products has shown any antimicrobial activities and the products have all had very different structures. In contrast, the example described here is clearly different. Firstly, the four copies of the α-peptide are essentially identical and all contain the highly conserved lipid II binding motif (Figure 2.1C),

strongly suggesting that the products are antimicrobial peptides. Secondly, the system contains two enzymes that have clearly defined roles, unlike the one enzyme in the cyanobacterial systems that has evolved for promiscuity. Thirdly, unlike the one cyanobacterial enzyme, which in order to be able to accommodate a highly diverse set of substrates is a very slow enzyme (89), the FlvM2 enzyme is very efficient in processing the eight FlvA2.x substrates that are diverse in sequence, but that have maintained a certain core set of rings at the C-terminus. All of these observations point at a two-component lantibiotic system where the diversification of the β-peptides has been beneficial in the specific growth environment of the producer organism.

In conclusion, this chapter presents the characterization of the ring topology and stereochemistry of nine different lanthipeptides encoded in the genome of *R. flavefaciens* FD-1. A subset of the modified Flvβ core peptides (Flvβ.b and Δ1-Flvβ.g) exhibited enhanced antibacterial activity when used in combination with Flvα.a, demonstrating that at least a subset of these peptides constitute two-component lantibiotic systems. Although the FlvA2 peptides display appreciable variation in their amino acid sequence, their post-translational modification is orchestrated by only one FlvM2 enzyme that is responsible for the modification of eight FlvA2 peptides with quite diverse sequences and final ring topologies. Elucidation of the features of the Flv system sets the stage for more in depth investigation of the biological roles of this curious set of lanthipeptides.

2.4 Materials and methods

2.4.1 General materials and methods

Molecular cloning reaction components were purchased from New England BioLabs (NEB) and included Phusion High-Fidelity DNA Polymerase, deoxynucleotide triphosphate mixture (dNTP) containing 10 mM of each deoxynucleotide, T4 DNA ligase. PCR and sequencing primers were ordered as custom oligonucleotide sequences from Integrated DNA Technologies. QIAquick PCR Purification Kit, QIAquick Gel Extraction Kit, and QIAprep Spin Miniprep Kit were purchased from Qiagen. For sequencing, the single pass primer extension sequencing service from ACGT, Inc. was used. Optima and LCMS grade acetonitrile (MeCN) was purchased from Fisher Scientific.

2.4.2 Cloning of *flv* genes

Chemically competent DH5α *E. coli* cells were used to maintain plasmids while BL21 (DE3) *E. coli* cells were used for the expression of constructs, which were assembled using a plasmid RSFDuet-1 backbone, except in the case of *flvT*. The *flvT* gene was incorporated into mcs1 of a plasmid ACYCDuet-1 backbone. Kanamycin was used at a final concentration of 50 µg/mL in both solid and liquid media. All liquid cultures were incubated at 37 °C with agitation at 200-220 revolutions per minute (rpm).

Polymerase chain reactions (PCRs) were run on a 20 µL scale and contained final concentrations of 1x Phusion HF Buffer (supplied with Phusion HF Polymerase), 3% v/v dimethylsulfoxide, 0.2 mM dNTPs, 0.5 µM of the forward and reverse primer, 56 ng of template DNA, and 10 U/mL of Phusion HF Polymerase. The PCRs were subjected to the following temperature (temp.) cycles: 30 s at 98 °C, 10 s at 98 °C, 30 s at the desired annealing temp., 30 s/kb at 72 °C, 5 min at 72 °C, and 4 °C for storage.

Table 2.2. Primers for cloning *flv* genes.

Primer name	Primer sequence 5′ to 3′	Annealing temp., °C
FlvAII.1_F Rum68_f (bamHI)	attactggatccgATGTCAGAGAAGAATATGGAAAAAG	51
FlvAII.1_R Rum68_b (hindIII)	attactaagcttTTACTTTCTTGGTAAAATCGG	48
FlvAII.2_F Rum69_f (bamHI)	attactggatccgATGGATAACAATACAAAATTACAGAAATTATAC	52
FlvAII.2_R Rum69_b (hindIII)	aatcctaagcttTCAAACTCTGCAAGAGTATGTGC	55
FlvAII.3_F 70_bam	attactggatccgATGGAAAATAAATTCGATATGGAAAAATTC	52
FlvAII.3_R 70_afl	attactcttaagTCAGAAACGGCAGGATGTTGTG	58
FlvAII.4_F Rum71_f (bamHI)	attactggatccgATGGATAACAATACTGAAAAATTCAACGAACTGGCAGC	58
FlvAII.4_R Rum71_b (hindIII)	attactcttaagTCAGAAACGGCAGGGCGGTG	58
FlvAII.5_F Rum72_f (bamHI)	attactggatccgATGAACAACAAGGAATTCAACATGGAACAG	58
FlvAII.5_R Rum72_b (hindIII)	attactaagcttTCAGAAACGGCAGGATGTTGTGCAAG	62
FlvAII.6_F Rum73_f (bamHI)	attactggatccgATGGAAAAGATGAACAATATAGCAGG	54
FlvAII.6_F Rum73_b (hindIII)	attactaagcttTCATAAAAGACGGCACTTCACTG	55
FlvAII.7_F Rum74_f (bamHI)	attactggatccgATGAACAACAATAACTTTGATATGGAAAAATTCAAGAAG	57

Table 2.2. (cont.)

FlvAII.7_R Rum74_b (hindIII)	attactaagcttTCAGAAGCGGCAGGAGTGTGTGCAAG	65
FlvAI.1_F Rum75/8_f (bamHI)	attactggatccgATGAACAAGAACCCAATCTACAG	53
FlvAI.1_R Rum75/8_b (hindIII)	attactaagcttTTACTTGTTGCAGGTCTTTGAG	53
FlvAII.8_F Rum79_f (bamHI)	attactggatccgATGGAAAGATACGGACATTTAGCAGG	57
FlvAII.8_R Rum79_b (hindIII)	attactaagcttTTATAATCGGCATGAATACGTGCAAG	56
FlvM1_F Rum81_f (asisI)	attactgcgatcgcGTGGATCAGCAGAACACACAGATAG	57
FlvM1_R Rum81_b (xho)	attactctcgagTCACGTCAACCAAAGATGATC	53
hisFlvM1_F(bamHI)	attactggatccgGTGGATCAGCAGAACACACAG	56
hisFlvM1_R (aflII)	attactcttaagTCACGTCAACCAAAGATGATCC	55
FlvM2_F Rum82_f (ndeI)	attactcatatgcaATGAAGCTGAATGAGATACATGATATCATAGACAACG	59
FlvM2_R Rum82_b (xho)	attactctcgagTCAGGATACAGCGAAAGCATATATC	54
hisFlvM2_F (bamHI)	attactggatccgATGAAGCTGAATGAGATACATGATATCATAGACAACG	59
hisFlvM2_R(aflII)	attactcttaagTCAGGATACAGCGAAAGCATATATCTCGCTATC	60
FlvT_F (ecoRI)	caccacagccaggatccgaattcaATGAATAGAAAATTAAGCGTACCAATGATAAAAC	67
FlvT_R (notI)	ctgttcgacttaagcattatgcggccgcTTAAGCAGACCTCAGTCTGTTGTAAAGTG	67
hisFlvT151_F (ndeI)	cctggtgccgcgcggcagccatattATGAATAGAAAATTAAGCGTACCAATGATAAAAC	67
hisFlvT151_R (bamHI)	ggccgcaagcttgtcgacggagctcgaattcggatccTTACACCCATATATTGGGGCCG	68

Aliquots of 5 µL of all PCRs were visualized using a 1% agarose gel and a Bio-Rad Molecular Imager XR+ Gel Doc Imaging System using the accompanying Quantity One 4.6.9 program. The remaining 15 µL of PCR sample was purified using a QIAquick PCR Purification Kit according to the manufacturer's protocol. Plasmids were digested using restriction endonucleases purchased from NEB according to the manufacturer's protocol and purified using a QIAquick Gel Extraction Kit also according to the manufacturer's protocol. Ligation reactions were run on a 5 µL scale using a 1:10 ratio of plasmid to insert using the manufacturer conditions described for T4 DNA ligase reactions. Following a 1:8 dilution, 10 µL of the ligation reaction was transformed into chemically competent *E. coli* DH5α cells. Transformants were screened for the desired construct by culturing in liquid media, isolating the DNA with a

QIAprep Spin Miniprep Kit and sequencing. For Quickchange mutagenesis, the PCR sample was subjected to DpnI digestion before PCR Purification. The resulting DNA was transformed into *E. coli* DH5α cells, and cells containing mutated plasmid were screened as described above.

2.4.3 Expression, IMAC, ZipTip, and MALDI-TOF-MS of Flv peptides

Electrocompetent *E. coli* BL21 (DE3) cells were transformed with the construct(s) to be expressed and plated on antibiotic containing media. Single colony transformants were then inoculated into starter liquid LB cultures containing 50 µg/mL of kanamycin. The starter culture was incubated overnight and expression cultures were then inoculated with the starter culture. The expression cultures were incubated until an optical density as monitored at 600 nm (OD_{600}) of 0.6-0.8 was achieved. A solution of isopropyl-β-D-galactopyranoside (1 M) was then added to the culture to achieve a final concentration of 1-0.5 mM. The expression culture was allowed to incubate for an additional 3-4 h. Cells from the expression culture were harvested by centrifugation at 5,000×*g*, 4 °C, for 15 min. The resulting pellet was stored at −80 °C until peptide purification was carried out.

Purification of the FlvA peptides was carried out as previously described with the following modifications to the protocol (90). Briefly, the cell pellets were re-solubilized in denaturing LanA Buffer 1 (6 M guanidine hydrochloride, 20 mM NaH_2PO_4, pH 7.5 at 25 °C, 500 mM sodium chloride, 0.5 mM imidazole) and lysed via sonication. After centrifugation of the lysate, the supernatant was clarified by filtration through a 0.45 µm membrane and subjected to IMAC. Desalting of the peptides in the eluent from IMAC was achieved through the use of Vydac BioSelect Solid Phase Extraction (SPE) Columns purchased from Grace (cat. no. 214SPE3000). The columns were first wetted with 2 column volumes (CV) of 100% acetonitrile (ACN) then equilibrated with 2 CV of 99.9% water, 0.1% trifluoroacetic acid (TFA). The eluent from IMAC was passed through an SPE column and the column bed was washed with 4 CV of the equilibration solution. Desalted peptide was eluted from the column using 80% ACN, 19.9% water, 0.1% TFA. This solution was flash frozen and dried *in vacuo* to obtain a fluffy white solid.

MALDI-TOF-MS analysis of desalted samples was completed using a Bruker UltrafleXtreme MALDI-TOFTOF-MS maintained in the UIUC School of Chemical Sciences Mass Spectrometry Laboratory. The Bruker flexControl and flexAnalysis programs were used

for data acquisition and analysis, respectively. The matrix suppression mode was set to deflection and the data was not subjected to processing. The instrument was calibrated before each use with Bruker Peptide Calibration Standard II or Protein Calibration Standard (Bruker Daltonics, Billerica, MA) prepared in the same way as the sample. For sample preparation, 0.5 µL of desalted sample was spotted on a Bruker MTP 384 polished steel target plate with an equal volume of sinapic acid matrix (80% ACN, 19.9% water, 0.1% TFA, sinapic acid added until saturation).

Proteolysis reactions ranged in volume from 20 µL to 20 mL and contained final concentrations of 250 mM Tris-HCl, pH 7.5, 2 mg/mL of peptide, and 0.2-1 µg/mL of protease. After approximately 1 h incubation at room temp., the reactions were desalted by ZipTip, as per the manufacturer's instruction (EMD Millipore). The content of the ZipTip was directly eluted onto a Bruker MTP 384 polished steel target plate with 1 µL of sinapic acid matrix. The sample was analyzed by MALDI-TOF-MS for the complete consumption of full length substrate peptide. The remainder of the peptide in the proteolysis reaction was desalted by SPE.

2.4.4 Expression and IMAC of FlvM and FlvT proteins

Expression of the FlvM and T proteins were carried out as described for the FlvA peptides, but instead of incubating at 37 °C after induction, the cells were incubated at 18 °C for 18-22 h. After incubation, the cells were harvested in the same manner as described for peptide expressing cultures. Purification of the FlvM and T proteins was carried out as previously described with the following modifications to the protocols (90, 91). The sample was maintained either on ice or at 4 °C for the duration of the protocol. Briefly, the cell pellets were re-solubilized in protein Start Buffer (LanM Start Buffer: 20 mM Tris, pH 7.6, 500 mM NaCl, 10% glycerol; LanT Start Buffer: 50 mM Na_2HPO_4, pH 7.5, 500 mM NaCl) and lysed using an Avestin EmulsiFlex-C3. After centrifugation of the lysate, the supernatant was clarified by filtration through a 0.45 µm membrane and subjected to IMAC using fast performance liquid chromatography (FPLC, Amersham Biosciences/GE Healthcare ÄKTA system) to provide a continuously increasing concentration of imidazole. Buffer exchange of the protein-containing eluent fractions from IMAC was achieved through the use of PD-10 size exclusion chromatography (GE Healthcare) and protein Final Buffer (LanM Final Buffer: 20 mM Tris, pH 7.6, 500 mM KCl, 10% glycerol; LanT Final Buffer: 50 mM Na_2HPO_4, pH 7.5, 500 mM NaCl,

10% glycerol). The concentration of protein was determined by monitoring the absorbance at 280 nM, and the protein solution was flash frozen and stored at −80 °C until in vitro assays were carried out.

2.4.5 IAA assays of FlvA peptides

The LanM in vitro and IAA assay reaction conditions were adapted from previous reports. In vitro reactions were setup using final concentrations of 200 mM HEPES, pH 8, 10 mM TCEP, 10 mM magnesium chloride, 5 mM adenosine triphosphate, 0.5-1 mg/mL of peptide, and 10-30 µg/mL of FlvM enzyme. Reactions ranged from 20 µL to 20 mL in volume. IAA assays were setup on a 20 µL volume using final concentrations of 250 mM HEPES, pH 9, 13 mM TCEP, 0.25 mg/mL of peptide, and 25 mM IAA. Reactions were allowed to incubate at room temp. for 1 h. For both types of reactions, reaction progress was monitored by ZipTip and MALDI-TOF-MS analysis as described above. In the case of the LanM reactions, the reactions were allowed to incubate at room temp. for additional time until the starting material peptide was no longer observed by MALDI-TOF-MS.

2.4.6 Preparation and MALDI-TOF-MS of FlvT co-expression samples

The pellet was removed from cultures obtained after the co-expression of FlvM2, A2.g and T by centrifugation at 11,900 ×*g*. The supernatant was then decanted and subjected to ZipTip and MALDI-TOF-MS analysis as stated above.

2.4.7 HPLC of FlvA peptides

Analytical scale RP-HPLC was completed using an Agilent 1260 Infinity equipped with a Phenomenex Luna column (10 µm, C18(2), 100 Å, 250×4.6 mm) and managed using Agilent Instrument 1 (Online) software. Semi-preparative scale RP-HPLC was performed using a Shimadzu Prominence equipped with a Phenomenex Luna column (10 µm, C18(2), 100 Å, 250×10.0 mm), managed using the Shimadzu program LC Real Time Analysis. A solvent system of A: 99.9% water, 0.1% TFA, B: 80% MeCN, 19.9% water, and 0.1% TFA was used. The gradient consisted of 0-5 min 2% B, 5-50 min 0-100% B, 50-60 min 2-100% B, and 60-65 min 100-2% B was used to purify the peptides. The core and GluC fragments of the FlvA peptides typically eluted between 30-40 min of the gradient. The resulting fractions of interest were

determined by MALDI-TOF-MS analysis, pooled, flash frozen and dried *in vacuo,* which yielded fluffy white solids.

2.4.8 Proteolysis of FlvA peptides

GluC was purchased from Roche Diagnostics (cat. no. 10791156001) and aminopeptidase was purchased from Sigma-Aldrich (cat. no. A8200-100UN). Both proteases were reconstituted using Milli-Q water to a final concentration of 10 mg/mL and 1 U/mL, respectively. Proteolysis reactions were carried out in 50-100 mM Tris-HCl, pH 8.0 with final concentrations of 1 mg/mL peptide and 0.04 mg/mL or 0.02 U/mL of protease. Reactions ranged in scale from 20 µL to 40 mL, and were incubated at room temp. for 12-18 h and analyzed by MALDI-TOF-MS. Proteolysis reactions were subjected to desalting and the peptide product was purified using RP-HPLC and dried as previously described.

2.4.9 Hydrolysis, derivatization, and GC/MS analysis of FlvA peptides

Peptide hydrolysis and amino acid derivatization was carried out as previously described (63). Briefly, the peptide was hydrolyzed in a sealed tube in 3 mL of 6 M HCl at 110 °C for 8-12 h. The hydrolysate was dried *in vacuo* and combined with a mixture of acetyl chloride that had been added drop-wise to 1.5 mL of methanol (MeOH). This reaction was refluxed at 110 °C for 45 min and then dried *in vacuo*. Then 3 mL of dichloromethane was added to the resulting residue, followed by the addition at 0 °C of 1 mL of pentafluoropropionic anhydride. The reaction was refluxed for 15 min at 110 °C, dried *in vacuo*, and stored at −80 °C until GC/MS analysis was performed. Prior to analysis, the sample was re-dissolved in 100 µL of MeOH. Particulates were removed by centrifugation at 23,700×*g* for 4 min, and the resulting supernatant was transferred into a glass insert and vial compatible with the instrument.

GC/MS analysis was performed using an Agilent Hewlett Packard 5973 mass spectrometer equipped with a CP-Chirasil-L Val column (25 m, 0.25 mm, 0.12 µm, 7 inch cage; Agilent Technologies). The initial oven temp. was 160 °C for 5 min followed by a gradient over 1 min to a final temp. of 180 °C, and then this temp. was maintained for 10 min. Sample was introduced into the instrument via a split (1:5) 2-5 µL injection with a flow rate of 2.0 mL/min helium gas. The initial inlet temp. was 190 °C, the MSD transfer line heater temp. was 185 °C, the MS Quad analyzer temp. was 150 °C, and the ion source temp was 230 °C. Scan and selected

ion monitoring at 365 Da and 379 Da data was acquired. To further confirm the identity of the species observed, authentic (Me)Lan standards of known stereochemistry were added to the samples and the resulting mixture was analyzed. Care was taken to dilute the samples or standards such that signal arising from (Me)Lan standard was comparable in intensity to that arising from the sample.

2.4.10 Tandem MS analysis of FlvA peptides

A Waters Synapt ESI-QTOF coupled to an Acquity ultra high performance liquid chromatography (UPLC) system was used for tandem mass spectrometry analysis. The UPLC was equipped with an Acquity UPLC BEH column (C8 1.7 µm 1.0×100 mm). The UPLC solvents used were A: 99.9% water, 0.1% formic acid (FA) and B: 99.9% MeCN, 0.1% FA. A gradient of 0-2 min 2% B, 2-12 min 2-98% B, 12-15 min 98% B, was employed. Tuning, calibration, and data acquisition was accomplished using the Waters MassLynx V4.1 program. Prior to sample analysis, the instrument was tuned and calibrated using a 0.1% phosphoric acid solution. All samples were run using [Glu1]-Fibrinopeptide B human (Sigma-Aldrich) as an external calibrant. Dried samples obtained after RP-HPLC were prepared for analysis by re-dissolving in Milli-Q water and transferring to a glass vial compatible with the UPLC system. Prior to tandem mass spectrometry, parent masses were verified by ms1 to be within 5 ppm of the calculated mass. Collision induced dissociation was used with a collision energy ramp ranging from 15-55 eV. Data analysis was completed using MassLynx V4.1 and the BioLynx Protein/Peptide editor.

2.4.11 Culturing the anaerobic bacterium *R. flavefaciens* FD-1

Liquid media was used to culture *R. flavefaciens* FD-1 (4.0 g of cellobiose, 2.0 g Bacto-tryptone, 50 mL mineral solution 1, 50 mL mineral solution 2, 10 mL volatile fatty acid solution in 820 mL of distilled water. After autoclaving but prior to inoculation 50 mL of an 8% sodium carbonate solution and 20 mL of a 1.25% cysteine-sulfide solution was added.

Mineral solution 1 consisted of 3.0 g potassium phosphate monobasic in 500 mL of distilled water; mineral solution 2 consisted of 3.0 g potassium phosphate monobasic, 6.0 g of sodium chloride, 0.6 g of magnesium sulfate, 0.6 g of calcium chloride dihyrdate, 6.0 g of ammonium sulfate in 500 mL of distilled water. The volatile fatty acid solution consisted of 6.85

mL of acetic acid, 3.0 mL of propionic acid, 1.84 mL of butyric acid, 0.55 mL of 2-methylbutyric acid, 0.47 mL of isobutyric acid, 0.55 mL of valeric acid, 0.55 mL of isovaleric acid, adjusted to pH 7.5 using sodium hydroxide and a final volume of 500 mL with distilled water.

Culturing procedures were performed using anaerobic technique and as previously described (19). *R. flavefaciens* FD-1 was inoculated directly into liquid media from revived glycerol stocks. Inoculated liquid cultures were incubated at 37 °C for 36-48 h. Subsequent steps were performed under aerobic conditions. The cells were pelleted by centrifugation at 11,900 ×*g*, 4 °C, for 10 min, and the resulting supernatant was subjected to ZipTip and MALDI-TOF-MS analysis for the presence of peptides. For solid medium co-culturing studies, 2 µL of each of the *R. flavefaciens* FD-1 and *R. albus* 7 liquid cultures was inoculated in close proximity on solid media. The plate was incubated for 24 h before colony MALDI-TOF-MS analysis. The samples were prepared by analyzing a small quantity of the proximal and distal (to the *R. albus* 7 colony) portion of the *R. flavefaciens* FD-1 colony and placing it on a Bruker MTP 384 polished steel target plate. One microliter of sinapic acid matrix was dispensed irectly on top of the sample and the sample was allowed to dry.

Aerobic agar diffusion assay media components were purchased from Becton, Dickinson and Company. Solid media was prepared with a 1.5% w/v final concentration of agar purchased from Fisher Scientific (cat. no. BP1423-500). All procedures were performed aseptically. Organisms were maintained as 40% glycerol stocks at −80 °C. The glycerol stocks were revived by first streaking onto solid media and incubating at 37 °C for 16-20 h. A single colony of the organism was transferred into 5 mL of liquid media and incubated at 37 °C, 200-220 rpm for 12-16 h. Agar diffusion assay plates were prepared by seeding 20 mL of molten solid media (~42 °C) with 50 µL of the overnight liquid culture. Following thorough mixing, the seeded solution was poured into a petri dish and allowed to solidify at room temp. Samples to be assayed were dissolved in DMSO and 1 µL aliquots were spotted directly onto the solidified media. When two peptide solutions were combined, only 0.5 µL aliquots of each solution were spotted in a single area. The Δ2-FlvA1.a peptide was assayed at a concentration of 1 mM while the FlvA2.x peptides were assayed at a concentration of 0.5 mM. Nisin was assayed at a concentration of 100 µM and 0.2 µL of the solution was spotted on the plate. Once all spots were dry, the petri dish

was incubated at 37 °C for 16-20 h prior to visualization using a Bio-Rad Molecular Imager XR+ Gel Doc Imaging System and the accompanying Quantity One 4.6.9 program.

2.5 References

1. Garrett, W.S. (2015). Cancer and the microbiota. *Science.* 348, 80-86.

2. Wlodarska, M., Kostic, A.D., and Xavier, R.J. (2015). An integrative view of microbiome-host interactions in inflammatory bowel diseases. *Cell Host Microbe.* 17, 577-591.

3. Yurkovetskiy, L.A., Pickard, J.M., and Chervonsky, A.V. (2015). Microbiota and autoimmunity: exploring new avenues. *Cell Host Microbe.* 17, 548-552.

4. Hooper, L.V., Littman, D.R., and Macpherson, A.J. (2012). Interactions between the microbiota and the immune system. *Science.* 336, 1268-1273.

5. Maynard, C.L., Elson, C.O., Hatton, R.D., and Weaver, C.T. (2012). Reciprocal interactions of the intestinal microbiota and immune system. *Nature.* 489, 231-241.

6. Tremaroli, V., and Backhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. *Nature.* 489, 242-249.

7. Nicholson, J.K., Holmes, E., and Wilson, I.D. (2005). Gut microorganisms, mammalian metabolism and personalized health care. *Nat. Rev. Microbiol.* 3, 431-438.

8. Vogt, S.L., Peña-Díaz, J., and Finlay, B.B. (2015). Chemical communication in the gut: effects of microbiota-generated metabolites on gastrointestinal bacterial pathogens. *Anaerobe.* 34, 106-115.

9. Hungate, R.E. (1966). The rumen and its microbes (New York, N.Y.: Academic Press).

10. White, B.A., Lamed, R., Bayer, E.A., and Flint, H.J. (2014). Biomass utilization by gut microbiomes. *Annu. Rev. Microbiol.* 68, 279-296.

11. Weimer, P.J. (2015). Redundancy, resilience, and host specificity of the ruminal microbiota: implications for engineering improved ruminal fermentations. *Front. Microbiol.* 6, 296.

12. Marcille, F., Gomez, A., Joubert, P., Ladiré, M., Veau, G., Clara, A., Gavini, F., Willems, A., and Fons, M. (2002). Distribution of genes encoding the trypsin-dependent lantibiotic ruminococcin A among bacteria isolated from human fecal microbiota. *Appl. Environ. Microbiol.* 68, 3424-3431.

13. Chen, J., Stevenson, D.M., and Weimer, P.J. (2004). Albusin B, a bacteriocin from the ruminal bacterium *Ruminococcus albus* 7 that inhibits growth of *Ruminococcus flavefaciens*. *Appl. Environ. Microbiol.* 70, 3167-3170.

14. Dabard, J., Bridonneau, C., Phillipe, C., Anglade, P., Molle, D., Nardi, M., Ladire, M., Girardin, H., Marcille, F., Gomez, A., and Fons, M. (2001). Ruminococcin A, a new lantibiotic produced by a *Ruminococcus gnavus* strain isolated from human feces. *Appl. Environ. Microbiol.* 67, 4111-4118.

15. Pujol, A., Crost, E.H., Simon, G., Barbe, V., Vallenet, D., Gomez, A., and Fons, M. (2011). Characterization and distribution of the gene cluster encoding RumC, an anti-*Clostridium perfringens* bacteriocin produced in the gut. *FEMS Microbiol. Ecol.* 78, 405-415.

16. Gomez, A., Ladire, M., Marcille, F., Nardi, M., and Fons, M. (2002). Characterization of ISRgn1, a novel insertion sequence of the IS3 family isolated from a bacteriocin-negative mutant of *Ruminococcus gnavus* E1. *Appl. Environ. Microbiol.* 68, 4136-4139.

17. Russell, J.B., and Mantovani, H.C. (2002). The bacteriocins of ruminal bacteria and their potential as an alternative to antibiotics. *J. Mol. Microbiol. Biotechnol.* 4, 347-355.

18. Shi, Y., and Weimer, P.J. (1996). Utilization of individual cellodextrins by three predominant ruminal cellulolytic bacteria. *Appl. Environ. Microbiol.* 62, 1084-1088.

19. Berg Miller, M.E., Antonopoulos, D.A., Rincon, M.T., Band, M., Bari, A., Akraiko, T., Hernandez, A., Thimmapuram, J., Henrissat, B., Coutinho, P.M., Borovok, I., Jindou, S., Lamed, R., Flint, H.J., Bayer, E.A., and White, B.A. (2009). Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of *Ruminococcus flavefaciens* FD-1. *PLoS One.* 4, e6650.

20. Lubelski, J., Rink, R., Khusainov, R., Moll, G.N., and Kuipers, O.P. (2008). Biosynthesis, immunity, regulation, mode of action and engineering of the model lantibiotic nisin. *Cell. Mol. Life. Sci.* 65, 455-476.

21. Letzel, A.C., Pidot, S.J., and Hertweck, C. (2014). Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. *BMC Genomics.* 15, 983.

22. Crost, E.H., Ajandouz, E.H., Villard, C., Geraert, P.A., Puigserver, A., and Fons, M. (2011). Ruminococcin C, a new anti-*Clostridium perfringens* bacteriocin produced in the gut by the commensal bacterium *Ruminococcus gnavus* E1. *Biochimie.* 93, 1487-1494.

23. Kalmokoff, M.L., and Teather, R.M. (1997). Isolation and characterization of a bacteriocin (Butyrivibriocin AR10) from the ruminal anaerobe *Butyrivibrio fibrisolvens* AR10: evidence in support of the widespread occurrence of bacteriocin-like activity among ruminal isolates of *B. fibrisolvens*. *Appl. Environ. Microbiol.* 63, 394-402.

24. Li, B., Sher, D., Kelly, L., Shi, Y., Huang, K., Knerr, P.J., Joewono, I., Rusch, D., Chisholm, S.W., and van der Donk, W.A. (2010). Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10430-10435.

25. Zhang, Q., Yang, X., Wang, H., and van der Donk, W.A. (2014). High divergence of the precursor peptides in combinatorial lanthipeptide biosynthesis. *ACS Chem. Biol.* 9, 2686-2694.

26. McAuliffe, O., Hill, C., and Ross, R.P. (2000). Each peptide of the two-component lantibiotic lacticin 3147 requires a separate modification enzyme for activity. *Microbiology.* 146, 2147-2154.

27. McClerren, A.L., Cooper, L.E., Quan, C., Thomas, P.M., Kelleher, N.L., and van der Donk, W.A. (2006). Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. *Proc. Natl. Acad. Sci. U.S.A.* 103, 17243-17248.

28. Begley, M., Cotter, P.D., Hill, C., and Ross, R.P. (2009). Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins. *Appl. Environ. Microbiol.* 75, 5451-5460.

29. Oman, T.J., and van der Donk, W.A. (2009). Insights into the mode of action of the two-peptide lantibiotic haloduracin. *ACS Chem. Biol.* 4, 865-874.

30. McAuliffe, O., Ryan, M.P., Ross, R.P., Hill, C., Breeuwer, P., and Abee, T. (1998). Lacticin 3147, a broad-spectrum bacteriocin which selectively dissipates the membrane potential. *Appl. Environ. Microbiol.* 64, 439-445.

31. Wiedemann, I., Bottiger, T., Bonelli, R.R., Wiese, A., Hagge, S.O., Gutsmann, T., Seydel, U., Deegan, L., Hill, C., Ross, P., and Sahl, H.G. (2006). The mode of action of the lantibiotic lacticin 3147-a complex mechanism involving specific interaction of two peptides and the cell wall precursor lipid II. *Mol. Microbiol.* 61, 285-296.

32. Oman, T.J., Lupoli, T.J., Wang, T.S., Kahne, D., Walker, S., and van der Donk, W.A. (2011). Haloduracin alpha binds the peptidoglycan precursor lipid II with 2:1 stoichiometry. *J. Am. Chem. Soc.* 133, 17544-17547.

33. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410.

34. Dischinger, J., Josten, M., Szekat, C., Sahl, H.G., and Bierbaum, G. (2009). Production of the novel two-peptide lantibiotic lichenicidin by *Bacillus licheniformis* DSM 13. *PLoS One.* 4, e6788.

35. Shenkarev, Z.O., Finkina, E.I., Nurmukhamedova, E.K., Balandin, S.V., Mineev, K.S., Nadezhdin, K.D., Yakimenko, Z.A., Tagaev, A.A., Temirov, Y.V., Arseniev, A.S., and Ovchinnikova, T.V. (2010). Isolation, structure elucidation, and synergistic antibacterial

activity of a novel two-component lantibiotic lichenicidin from *Bacillus licheniformis* VK21. *Biochemistry.* 49, 6462-6472.

36. Caetano, T., Krawczyk, J.M., Mösker, E., Süssmuth, R.D., and Mendo, S. (2011). Heterologous expression, biosynthesis, and mutagenesis of type II lantibiotics from *Bacillus licheniformis* in *Escherichia coli*. *Chem. Biol.* 18, 90-100.

37. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., and Higgins, D.G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539.

38. Szekat, C., Jack, R.W., Skutlarek, D., Farber, H., and Bierbaum, G. (2003). Construction of an expression system for site-directed mutagenesis of the lantibiotic mersacidin. *Appl. Environ. Microbiol.* 69, 3777-3783.

39. Dufour, A., Hindré, T., Haras, D., and Le Pennec, J.P. (2007). The biology of the lantibiotics of the lacticin 481 subgroup is coming of age. *FEMS Microbiol. Rev.* 31, 134-167.

40. Hsu, S.T., Breukink, E., Bierbaum, G., Sahl, H.G., de Kruijff, B., Kaptein, R., van Nuland, N.A., and Bonvin, A.M. (2003). NMR study of mersacidin and lipid II interaction in dodecylphosphocholine micelles. Conformational changes are a key to antimicrobial activity. *J. Biol. Chem.* 278, 13110-13117.

41. Islam, M.R., Nishie, M., Nagao, J., Zendo, T., Keller, S., Nakayama, J., Kohda, D., Sahl, H.G., and Sonomoto, K. (2012). Ring A of nukacin ISK-1: a lipid II-binding motif for type-A(II) lantibiotic. *J. Am. Chem. Soc.* 134, 3687-3690.

42. Martin, N.I., Sprules, T., Carpenter, M.R., Cotter, P.D., Hill, C., Ross, R.P., and Vederas, J.C. (2004). Structural characterization of lacticin 3147, a two-peptide lantibiotic with synergistic activity. *Biochemistry.* 43, 3049-3056.

43. Tang, W., and van der Donk, W.A. (2013). The sequence of the enterococcal cytolysin imparts unusual lanthionine stereochemistry. *Nat. Chem. Biol.* 9, 157-159.

44. Tang, W., Jiménez-Osés, G., Houk, K.N., and van der Donk, W.A. (2015). Substrate control in stereoselective lanthionine biosynthesis. *Nat. Chem.* 7, 57-64.

45. Lohans, C.T., Li, J.L., and Vederas, J.C. (2014). Structure and biosynthesis of carnolysin, a homologue of enterococcal cytolysin with D-amino acids. *J. Am. Chem. Soc.* 136, 13150-13153.

46. Traxler, M.F., Watrous, J.D., Alexandrov, T., Dorrestein, P.C., and Kolter, R. (2013). Interspecies interactions stimulate diversification of the *Streptomyces coelicolor* secreted metabolome. *Mbio.* 4, e00459-00413.

47. Yang, Y.L., Xu, Y.Q., Kersten, R.D., Liu, W.T., Meehan, M.J., Moore, B.S., Bandeira, N., and Dorrestein, P.C. (2011). Connecting chemotypes and phenotypes of cultured marine microbial assemblages by imaging mass spectrometry. *Angew. Chem. Int. Ed.* 50, 5839-5842.

48. Bode, H.B., and Müller, R. (2005). The impact of bacterial genomics on natural product research. *Angew. Chem., Int. Ed. Engl.* 44, 6828-6846.

49. Winter, J.M., Behnken, S., and Hertweck, C. (2011). Genomics-inspired discovery of natural products. *Curr. Opin. Chem. Biol.* 15, 22-31.

50. Shi, Y., Yang, X., Garg, N., and van der Donk, W.A. (2011). Production of lantipeptides in *Escherichia coli*. *J. Am. Chem. Soc.* 133, 2338-2341.

51. Lin, Y., Teng, K., Huan, L., and Zhong, J. (2011). Dissection of the bridging pattern of bovicin HJ50, a lantibiotic containing a characteristic disulfide bridge. *Microbiol. Res.* 166, 146-154.

52. Nagao, J., Harada, Y., Shioya, K., Aso, Y., Zendo, T., Nakayama, J., and Sonomoto, K. (2005). Lanthionine introduction into nukacin ISK-1 prepeptide by co-expression with modification enzyme NukM in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 336, 507-513.

53. Wang, J., Ma, H., Ge, X., Zhang, J., Teng, K., Sun, Z., and Zhong, J. (2014). Bovicin HJ50-like lantibiotics, a novel subgroup of lantibiotics featured by an indispensable disulfide bridge. *PLoS One.* 9, e97121.

54. Basi-Chipalu, S., Dischinger, J., Josten, M., Szekat, C., Zweynert, A., Sahl, H.G., and Bierbaum, G. (2015). Pseudomycoicidin, a class II lantibiotic from *Bacillus pseudomycoides*. *Appl. Environ. Microbiol.* 81, 3419-3429.

55. Garg, N., Tang, W., Goto, Y., Nair, S.K., and van der Donk, W.A. (2012). Lantibiotics from *Geobacillus thermodenitrificans*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5241-5246.

56. Ökesli, A., Cooper, L.E., Fogle, E.J., and van der Donk, W.A. (2011). Nine post-translational modifications during the biosynthesis of cinnamycin. *J. Am. Chem. Soc.* 133, 13753-13760.

57. Geoghegan, K.F., Dixon, H.B., Rosner, P.J., Hoth, L.R., Lanzetti, A.J., Borzilleri, K.A., Marr, E.S., Pezzullo, L.H., Martin, L.B., LeMotte, P.K., McColl, A.S., Kamath, A.V., and Stroh, J.G. (1999). Spontaneous alpha-N-6-phosphogluconoylation of a "His tag" in *Escherichia coli*: the cause of extra mass of 258 or 178 Da in fusion proteins. *Anal. Biochem.* 267, 169-184.

58. Rink, R., Kuipers, A., de Boef, E., Leenhouts, K.J., Driessen, A.J., Moll, G.N., and Kuipers, O.P. (2005). Lantibiotic structures as guidelines for the design of peptides that can be modified by lantibiotic enzymes. *Biochemistry.* 44, 8873-8882.

59. Thibodeaux, G.N., McClerren, A.L., Ma, Y., Gancayco, M.R., and van der Donk, W.A. (2015). Synergistic binding of the leader and core peptides by the lantibiotic synthetase HalM2. *ACS Chem. Biol.* 10, 970-977.

60. Yang, X., and van der Donk, W.A. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: new insights into the role of leader and core peptides during biosynthesis. *Chem. Eur. J.* 19, 7662-7677.

61. Uguen, P., Hindré, T., Didelot, S., Marty, C., Haras, D., Le Pennec, J.P., Vallee-Réhel, K., and Dufour, A. (2005). Maturation by LctT is required for biosynthesis of full-length lantibiotic lacticin 481. *Appl. Environ. Microbiol.* 71, 562-565.

62. Håvarstein, L.S., Diep, D.B., and Nes, I.F. (1995). A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Mol. Microbiol.* 16, 229-240.

63. Ross, A.C., Liu, H., Pattabiraman, V.R., and Vederas, J.C. (2010). Synthesis of the lantibiotic lactocin S using peptide cyclizations on solid phase. *J. Am. Chem. Soc.* 132, 462-463.

64. Kabuki, T., Uenishi, H., Seto, Y., Yoshioka, T., and Nakajima, H. (2009). A unique lantibiotic, thermophilin 1277, containing a disulfide bridge and two thioether bridges. *J. Appl. Microbiol.* 106, 853-862.

65. Lin, Y.H., Teng, K.L., Huan, L.D., and Zhong, J. (2011). Dissection of the bridging pattern of bovicin HJ50, a lantibiotic containing a characteristic disulfide bridge. *Microbiol. Res.* 166, 146-154.

66. Zhang, J., Feng, Y.G., Teng, K.L., Lin, Y.H., Gao, Y., Wang, J.F., and Zhong, J. (2014). Type All lantibiotic bovicin HJ50 with a rare disulfide bond: structure, structure-activity relationships and mode of action. *Biochem. J.* 461, 497-508.

67. Oman, T.J., Boettcher, J.M., Wang, H.A., Okalibe, X.N., and van der Donk, W.A. (2011). Sublancin is not a lantibiotic but an S-linked glycopeptide. *Nat. Chem. Biol.* 7, 78-80.

68. Tang, W., and van der Donk, W.A. (2012). Structural characterization of four prochlorosins: a novel class of lantipeptides produced by planktonic marine cyanobacteria. *Biochemistry.* 51, 4271-4279.

69. Plat, A., Kluskens, L.D., Kuipers, A., Rink, R., and Moll, G.N. (2011). Requirements of the engineered leader peptide of nisin for inducing modification, export, and cleavage. *Appl. Environ. Microbiol.* 77, 604-611.

70. Bindman, N.A., and van der Donk, W.A. (2013). A general method for fluorescent labeling of the N-termini of lanthipeptides and its application to visualize their cellular localization. *J. Am. Chem. Soc.* 135, 10362-10371.

71. Shi, Y., Bueno, A., and van der Donk, W.A. (2012). Heterologous production of the lantibiotic Ala(0)actagardine in *Escherichia coli*. *Chem. Commun.* 48, 10966-10968.

72. Majchrzykiewicz, J.A., Lubelski, J., Moll, G.N., Kuipers, A., Bijlsma, J.J., Kuipers, O.P., and Rink, R. (2010). Production of a class II two-component lantibiotic of *Streptococcus pneumoniae* using the class I nisin synthetic machinery and leader sequence. *Antimicrob. Agents Chemother.* 54, 1498-1505.

73. Velásquez, J.E., Zhang, X., and van der Donk, W.A. (2011). Biosynthesis of the antimicrobial peptide epilancin 15X and its unusual N-terminal lactate moiety. *Chem. Biol.* 18, 857-867.

74. D'Onofrio, A., Crawford, J.M., Stewart, E.J., Witt, K., Gavrish, E., Epstein, S., Clardy, J., and Lewis, K. (2010). Siderophores from neighboring organisms promote the growth of uncultured bacteria. *Chem. Biol.* 17, 254-264.

75. Singh, M., and Sareen, D. (2014). Novel LanT associated lantibiotic clusters identified by genome database mining. *PLoS One.* 9, e91352.

76. Odenyo, A.A., Mackie, R.I., Stahl, D.A., and White, B.A. (1994). The use of 16S rRNA-targeted oligonucleotide probes to study competition between ruminal fibrolytic bacteria: development of probes for *Ruminococcus* species and evidence for bacteriocin production. *Appl. Environ. Microbiol.* 60, 3697-3703.

77. Kalmokoff, M.L., Lu, D., Whitford, M.F., and Teather, R.M. (1999). Evidence for production of a new lantibiotic (butyrivibriocin OR79A) by the ruminal anaerobe *Butyrivibrio fibrisolvens* OR79: characterization of the structural gene encoding butyrivibriocin OR79A. *Appl. Environ. Microbiol.* 65, 2128-2135.

78. Gomez, A., Ladire, M., Marcille, F., and Fons, M. (2002). Trypsin mediates growth phase-dependent transcriptional regulation of genes involved in biosynthesis of ruminococcin A, a lantibiotic produced by a *Ruminococcus gnavus* strain from a human intestinal microbiota. *J. Bacteriol.* 184, 18-28.

79. Cooper, L.E., McClerren, A.L., Chary, A., and van der Donk, W.A. (2008). Structure-activity relationship studies of the two-component lantibiotic haloduracin. *Chem. Biol.* 15, 1035-1045.

80. Cotter, P.D., Deegan, L.H., Lawton, E.M., Draper, L.A., O'Connor, P.M., Hill, C., and Ross, R.P. (2006). Complete alanine scanning of the two-component lantibiotic lacticin 3147: generating a blueprint for rational drug design. *Mol. Microbiol.* 62, 735-747.

81. Backhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., and Gordon, J.I. (2005). Host-bacterial mutualism in the human intestine. *Science.* 307, 1915-1920.

82. Kim, M., Morrison, M., and Yu, Z.T. (2011). Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiol. Ecol.* 76, 49-63.

83. Dethlefsen, L., McFall-Ngai, M., and Relman, D.A. (2007). An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature.* 449, 811-818.

84. Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R., and Gordon, J.I. (2008). Evolution of mammals and their gut microbes. *Science.* 320, 1647-1651.

85. Wang, H., Oman, T.J., Zhang, R., De Gonzalo, C.V.G., Zhang, Q., and van der Donk, W.A. (2014). The glycosyltransferase involved in thurandacin biosynthesis catalyzes both O- and S-glycosylation. *J. Am. Chem. Soc.* 136, 84-87.

86. Scholz, R., Molohon, K.J., Nachtigall, J., Vater, J., Markley, A.L., Süssmuth, R.D., Mitchell, D.A., and Borriss, R. (2011). Plantazolicin, a novel microcin B17/streptolysin S-like natural product from *Bacillus amyloliquefaciens* FZB42. *J. Bacteriol.* 193, 215-224.

87. Azevedo, A.C., Bento, C.B., Ruiz, J.C., Queiroz, M.V., and Mantovani, H.C. (2015). Distribution and genetic diversity of bacteriocin gene clusters in rumen microbial genomes. *Appl. Environ. Microbiol.* 81, 7290-7304.

88. Li, B., Sher, D., Kelly, L., Shi, Y.X., Huang, K., Knerr, P.J., Joewono, I., Rusch, D., Chisholm, S.W., and van der Donk, W.A. (2010). Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10430-10435.

89. Thibodeaux, C.J., Ha, T., and van der Donk, W.A. (2014). A price to pay for relaxed substrate specificity: a comparative kinetic analysis of the class II lanthipeptide synthetases ProcM and HalM2. *J. Am. Chem. Soc.* 136, 17513-17529.

90. Li, B., Cooper, L.E., and van der Donk, W.A. (2009). Chapter 21. In vitro studies of lantibiotic biosynthesis. *Method. Enzymol.* 458, 533-558.

91. Furgerson Ihnken, L.A., Chatterjee, C., and van der Donk, W.A. (2008). In vitro reconstitution and substrate specificity of a lantibiotic protease. *Biochemistry.* 47, 7352-7363.

Chapter 3. Identification of venezuelin-like peptides from *Actinobacteria*[*]

3.1 Introduction

As discussed in Chapter 1, lanthipeptides are widely distributed in taxonomically distant species and display very diverse biological activities, ranging from antimicrobial to antiallodynic (1-3). Four classes of biosynthetic enzymes are known to catalyze lanthionine formation (4, 5). The four distinct classes of biosynthetic machinery reflect the functional importance of lanthionine scaffolds and a convergent evolution process to produce them. Notably, products generated by these four classes of lanthionine synthetases are not limited to lanthipeptides. For example, many lanthipeptide-like biosynthetic gene clusters encode precursor peptides that have no Cys residues, and their products are therefore not lanthipeptides (6).

Over the past 5 years, several studies have reported bioinformatics genome mining efforts based on the then-available genomes. These studies have typically focused on an enzyme-specific query, such as class I LanB and LanC enzymes (7), class II LanM proteins (8), or the bifunctional LanT proteins, which transport lanthipeptides and remove their leader sequences (9). Building on the recent natural product gene cluster family (GCF) project (10), this chapter presents a large-scale analysis of lanthipeptide-like biosynthetic gene clusters that is driven by similarities in entire clusters rather than specific enzymes. We focused the analysis on lanthipeptides from *Actinobacteria*, which are relatively understudied compared to family members from Firmicutes. *Actinobacteria* carry a wealth of natural product biosynthetic gene clusters whose products encompass highly diverse chemical structures.

This phylum has been the source and/or inspiration for the majority of pharmaceutically useful natural products (11). Although *Actinobacteria* constitute only a small proportion of known lanthipeptide producers, they encode a diverse set of post-translational modifications that have not been observed for lanthipeptides from other phyla (4) (e.g., lysinoalanine formation in cinnamycin (12), tryptophan chlorination and proline dihydroxylation in microbisporicin (13, 14) also called NAI-107 (15)), and glycosylation in NAI-112 (16). The unique modifications in

[*] Portions are reproduced with permission from 181. Zhang, Q., Doroghazi, J.R., Zhao, X., Walker, M.C., and van der Donk, W.A. (2015). Expanded natural product diversity revealed by analysis of lanthipeptide-like gene clusters in *Actinobacteria*. *Appl. Environ. Microbiol.* 81, 4339-4350. Studies were performed in collaboration with Dr. J. Doroghazi, Dr. Q. Zhang, and Dr. M. Walker. Dr. Doroghazi conceived of and constructed the GCF networks. Dr. Q. Zhang, and Dr. M. Walker analyzed all members of the GCFs. X. Zhao identified, isolated, and characterized the SapB and venezuelin peptides.

actinobacterial lanthipeptide biosynthesis suggest great potential for exploring unknown lanthipeptide chemical space in this phylum.

3.2 Results

3.2.1 SapB identification from *Actinobacteria*

Of the *Actinobacteria* screened, several strains yielded ions with masses that corresponded to putative core peptides from GCF31 (Figure 3.1), and this GCF corresponds to that of the morphogenic peptide SapB. Since the fragmentation pattern of the 2027 m/z ion from a member of GCF31, *Streptomyces* species S-1833, was nearly identical to that reported by Kodani and co-workers for SapB (Figure 3.2) (17), further characterization of the compound was not pursued. The ubiquity of SapB-like compounds is also evidenced in the observations of Dorrestein and co-workers (18).

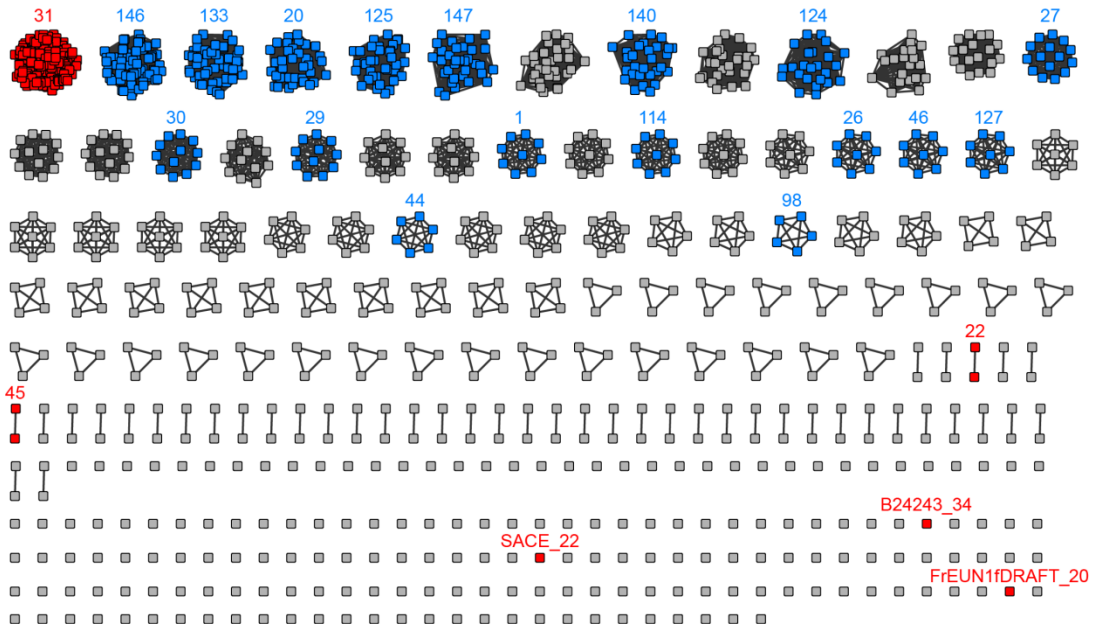

Figure 3.1. Lanthipeptide gene cluster family similarity network for *Actinobacteria*. Lanthipeptide gene cluster families discussed in (19) are shown in blue, and gene cluster families with previously characterized members or close analogs are shown in red. GCFs are labeled with their respective numbers whereas gene clusters not belonging to families (i.e. singletons) are labeled with their cluster identifier. GCFs in grey are not discussed in this chapter nor (19), but can be found on the website http://www.igb.illinois.edu/labs/metcalf/gcf/lant.html., together with their GCF numbers. To access the genes in the singleton clusters on the website, mouse over the singleton to show the GCF number. For instance, the very first singleton is Saci8_3. Then click Search at the top, and enter Saci8_3 in the

Figure 3.1. (cont.)

search field for Gene Cluster Search. Then submit and click on the Saci8_3 link that appears. Previously characterized lanthipeptides or their analogs that are found in GCF31: SapB and various peptins (class III lanthipeptides); GCF22: actagardine (class II); GCF45: microbisporicin (class I); B24243_34 from *Streptomyces africanus* B-24243 is similar to cinnamycin (72% identical LanAs, class II); SACE_22: erythreapeptin from *Saccharopolyspora erythraea* NRRL 2338 (class III); FrEUN1fDRAFT_20 from *Frankia* sp. EUN1f is similar to cinnamycin (54% identical LanAs; class II). Figure was made by Dr. J. Doroghazi.

In our mass spectrometric screen of *Actinobacteria*, correlation of a predicted gene cluster and observed ion in the absence of fragmentation data was complicated by the inability to accurately calculate a putative mass, which was due to uncertainty regarding the length of the leader peptides. Cleavage sites and the potential for post-translational modifications that did not resemble characterized ones result in substantial shifts in calculated masses. Although more automated tandem MS data acquisition could overcome the necessity of accurate parent mass calculation, low production or silent gene clusters would remain problematic.

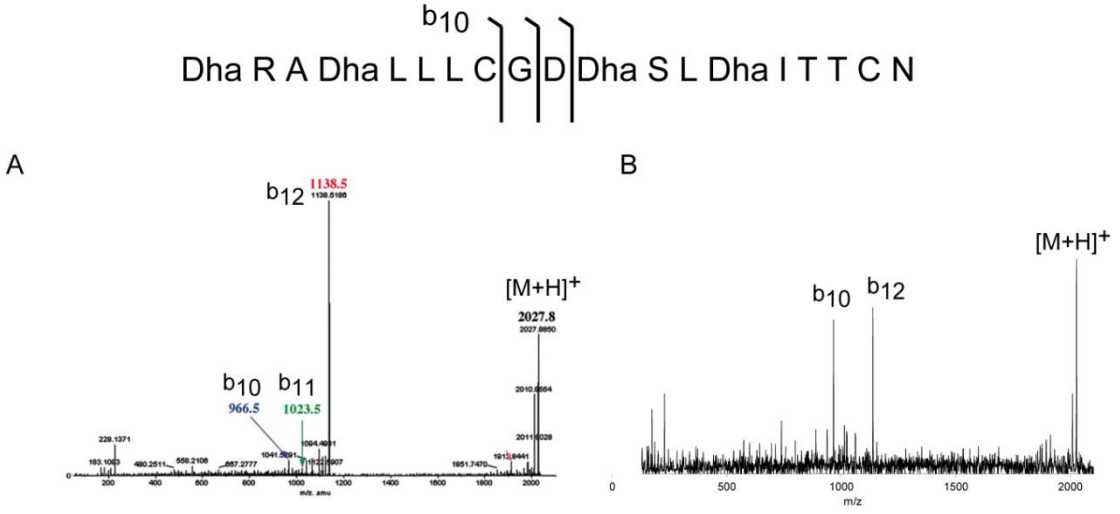

Figure 3.2. Fragmentation pattern of SapB. Tandem MS spectrum of (A) previously reported SapB. Panel A reproduced with permission from (17), copywrite (2004) National Academy of Sciences, U.S.A. (B) Fragmentation of SapB isolated from *S.* species S-1833.

However, the GCF organization allowed for indirect correlation of masses with a gene cluster shared among several *Actinobacteria*. This approach is illustrated by the set of masses in the 1700-2200 m/z range observed from a set of the *Actinobacteria* that share a class IV, venezuelin-like gene cluster (Figure 3.3A, B, E).

3.2.2 Identification and characterization of class IV lanthipeptides

At the time of this study, venezuelin had been obtained only via in vitro modification of the precursor peptide VenA by VenL followed by proteolytic removal of the leader peptide (20). However, production of venezuelin was detected neither in the wild-type strain *S. venezuelae* nor after attempted expression of the gene cluster in *S. lividans*, raising the question whether venezuelin is a genuine natural lanthipeptide. In our current analysis, we noted that GCF147 (33 members) (Figure 3.1) consists mostly of venezuelin-like gene clusters (Figure 3.3C), and the putative precursor peptide sequences are highly conserved (Figure 3.3D), suggesting the wide occurrence of venezuelin-like lanthipeptide gene clusters (although it was first discovered in *S. venezuelae*, here we collectively designate all these similar putative lanthipeptides venezuelins). The conserved genes in GCF147 include a *lanA* gene, a *lanL* gene, and two ABC transporter genes (Figure 3.3C), indicating that (Me)Lan formation is likely the only post-translational modification on LanA core peptides in this family. Three *Streptomyces* strains (*S. katrae* ISP5550, *S. lavendulae* subsp. lavendulae NRRL B-2508, and *S.* sp. strain NRRL B-2375) produced venezuelins in detectable amounts (Figure 3.3A, B, E). Venezuelins of different lengths were detected, similar to the observations with class III lanthipeptides (21-24). The variation in lengths of the observed venezuelins corresponded to the presence or absence of Phe28 and Ala29 (numbering based on VenA sequence) (Figure 3.3A, B, E), presumably arising from stepwise removal of (part of) the leader peptides.

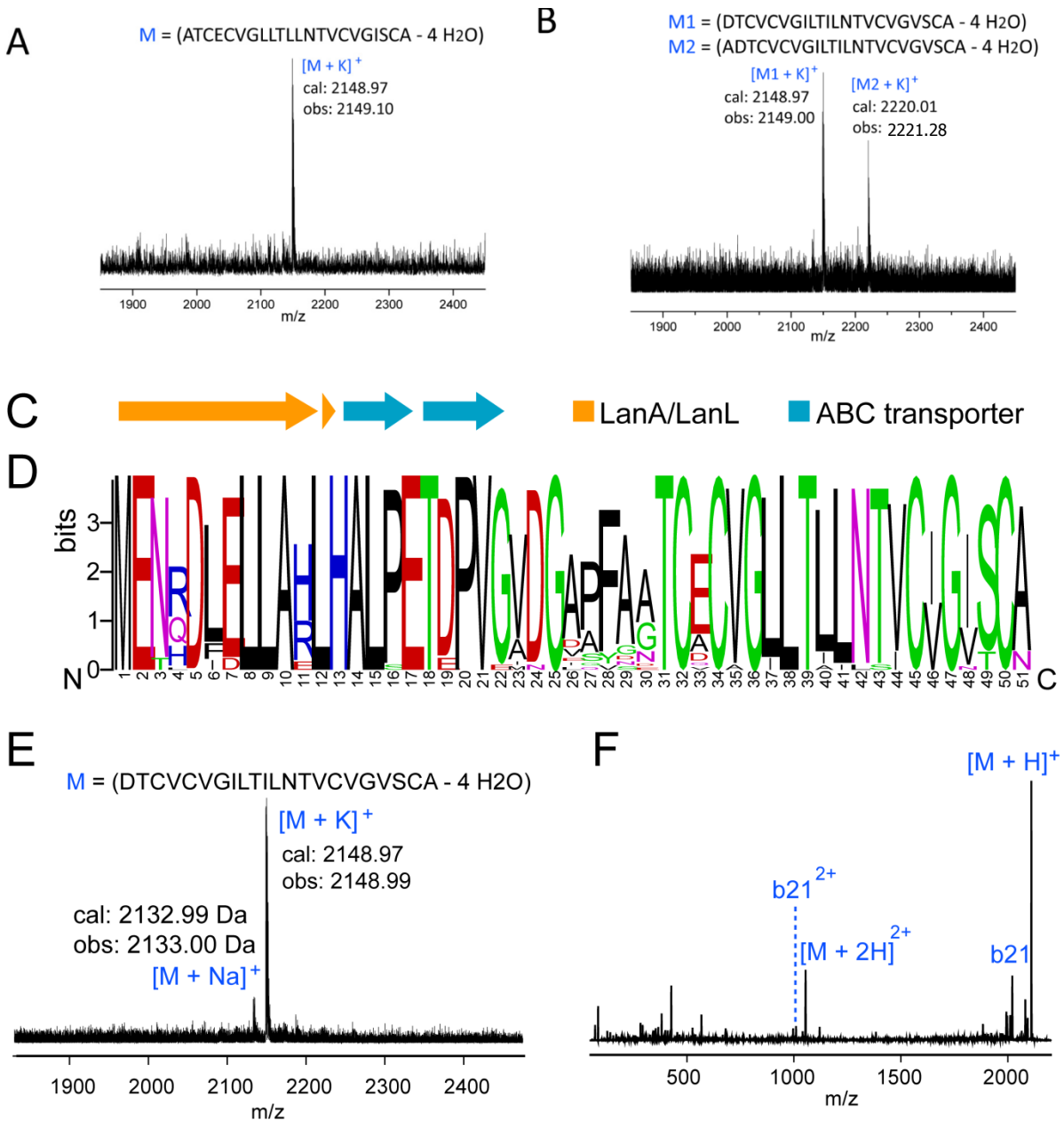

Figure 3.3. MALDI-TOF MS detection of venezuelins in *Actinobacteria*. Spectra from colony MS of (A) *S.* species B-2375, (B) *S. lavendulae* subspecies lavendulae B-2508, and (C) representative gene clusters of GCF147. (D) A logo depicting the GCF147 precursor peptide sequences. (E) MALDI-TOF MS analysis of venezuelin produced by *S. katrae* ISP5550. (F) ESI-MS-MS analysis of venezuelin produced by *S. katrae* ISP5550. The lack of fragmentation is consistent with a globular, overlapping ring topology.

The venezuelin from *S. katrae* ISP5550, which contained Ala29 at its N-terminus, was further analyzed by electrospray ionization (ESI)-MS-MS analysis (Figure 3.3F). The lack of any fragmentation showed that the peptide likely consists of overlapping (Me)Lan rings, consistent with the previously proposed ring topology of venezuelins based on extensive mutagenesis studies (20). Additional verification that venezuelin from *S. katrae* ISP5550 is indeed a lanthipeptide was provided by hydrolysis of the peptide and derivatization of the constituent amino acids as methyl esters and pentafluoropropionamides (25, 26). Chiral GC-MS analysis of the derivatized amino acids revealed the presence of (Me)Lan, supporting the designation of venezuelin as a lanthipeptide (Figure 3.4). The (Me)Lan derivatives all had the DL configuration.

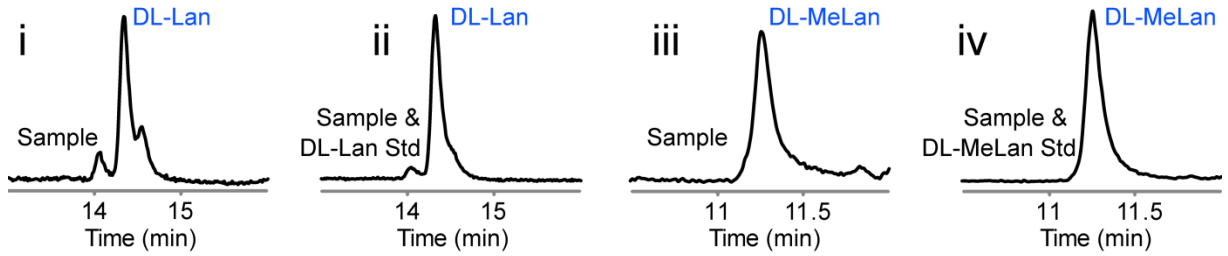

Figure 3.4. Chiral GC-MS analysis of hydrolyzed and derivatized venezuelin residues, which revealed the presence of DL-Lan and DL-MeLan. Trace i, sample showing the presence of DL-lanthionine obtained from venezuelin; trace ii, sample spiked with DL-Lan standard; trace iii, sample showing the presence of DL-MeLan obtained from venezuelin; trace iv, sample spiked with DL-MeLan standard.

Our analysis validates that venezuelins are genuine natural products, and based on the frequency of their biosynthetic genes (Figure 3.1), they may be produced by many *Actinobacteria*. This observation also further supports the previous suggestion that identification of the products of silent clusters may be accomplished by screening several strains that contain the GCF (10).

3.3 Discussion

Lanthipeptides are an intriguing class of natural products not only because of their potential medicinal values but also because of their wide occurrence and suitability for genome mining and heterologous expression studies. The identification of venezuelin from *S. katrae*

ISP5550, *S. lavendulae* subsp. lavendulae NRRL B-2508, and *S.* sp. strain NRRL B-2375 was greatly facilitated by the GCFs. Also, the organization allowed for intuitive appreciation of the biosynthetic diversity, an aspect of bioinformatics discussed in Chapter 1. The study presented herein highlights how bioinformatic organization of genome information complements and accelerates discovery efforts.

The low frequency of masses corresponding to putative products can be attributed to the recalcitrant or silent nature of biosynthetic gene clusters, which was observed in the original report of venezuelin (20). Most of the class III lanthipeptides reported thus far from *Actinobacteria* have no or weak antimicrobial activities, and quite a few have anti-allodynic/antinociceptive activity (16, 21) or morphogenetic activities (27). Since the venezuelins did not display antimicrobial activity in a previous study (20), class IV lanthipeptides may also have activities that would have eluded antimicrobial screens. Some evidence for this was presented upon isolation of larger quantities of another class IV lanthipeptide, streptocollin. Although this peptide did not exhibit antimicrobial, antiviral, or morphogenetic activity, it did inhibit protein tyrosine phosphatase 1B, which is involved in signaling pathways (28).

3.4 Materials and methods

3.4.1 Colony MS

A small portion (approximately 1x1 mm) of colony was transferred onto a Bruker MTP 384 polished steel target plate. On top of the colony, 2 µL of sinapic acid matrix was overlaid and allowed to air dry. Sinapic acid matrix was prepared by saturating an 80 acetonitrile (MeCN): 20 Milli-Q water solution with sinapic acid (Sigma-Aldrich). MALDI-TOF MS instrumentation, software, and calibration were the same as described in Chapter 2, Materials and methods section.

3.4.2 Solid cultures of *S. katrae* ISP5550

DMSO stocks of *S. katrae* ISP5550 were prepared by mixing 600 µL of a *S. katrae* ISP5550 liquid culture with an equal volume of sterilized 20% dimethyl sulfoxide (DMSO). The resulting solution was mixed, flash frozen, and stored at −80 °C. *S. katrae* ISP5550 was inoculated from DMSO stocks onto solid ATCC172 media (20 g soluble starch, 5 g yeast extract,

5 g N-Z amine type A, 1 g calcium carbonate, 15 g agar, 1 L of Milli-Q water, after autoclaving the media, sterilized glucose was added to a 1% final concentration) and incubated at 30 °C for 4-5 days. The presence of venezuelins was detected using colony MALDI-TOF-MS as described in the preceding paragraph.

3.4.3 Liquid cultures of *S. katrae* ISP5550

Upon detection of venezuelins by colony MALDI-TOF-MS from solid cultures of *S. katrae* ISP5550, liquid cultures were analyzed for the production of venezuelins. A single colony was transferred into 50-100 mL of liquid ATCC172 media, prepared in the same way as the solid media with the omission of agar. The liquid culture was incubated with shaking at 30 °C for 12 days. Supernatant from the liquid cultures was obtained by centrifuging at 4,650 ×g for 15 min. The hydrophobic components of the supernatant were desalted by ZipTip (EMD Millipore) according to the manufacturer's instructions. The ZipTip contents were directly eluted onto a Bruker MTP 384 polished steel target plate on a spot containing 1 µL of sinapic acid matrix. Analysis by MALDI-TOF-MS was performed as described above.

In preparation for ESI MS, peptide components of the supernatant were precipitated using ammonium sulfate. A 60% final concentration of ammonium sulfate was achieved with the addition of 30 g of ammonium sulfate to 50 mL of supernatant. The solution was allowed to incubate at room temperature for 30 min. The resulting precipitate was pelleted by centrifugation at 22,789 ×g for 10 min, and the supernatant was decanted. The pellet was re-dissolved in Milli-Q water and subjected to ESI-MS analysis using an Acquity ultra performance liquid chromatography (UPLC) system coupled to a Waters Synapt ESI-QTOF. The UPLC was equipped with an Acquity UPLC BEH C8 1.7 µm 1.0×100 mm column. Sample was introduced into the column via an Acquity autoinjector and subjected to a gradient of 2% B from 0-2 min, 2%-98% B from 2-12 min, and 98% B from 12-15 min (A was 0.1% formic acid in water and B was 0.1% formic acid in MeCN). The Waters MassLynx V4.1 program was used to tune, calibrate and acquire data. [Glu1]-Fibrinopeptide B human (Sigma-Aldrich) was used as an external calibrant and the mass of venezuelin was verified to be within 5 ppm of its calculated mass. Collision induced dissociation was achieved using a series of energy ramps ranging from 30-60 eV.

3.4.4 Isolation of venezuelins

S. katrae ISP5550 liquid culture supernatant was subjected to pressure-based sample concentration rather than ammonium sulfate precipitation. After centrifugation, 40 mL of culture supernatant was further clarified by filtration through a 0.22 µm nylon membrane and then loaded into an Amicon stirred cell equipped with a 1 kDa nominal molecular weight limit regenerated cellulose disc (EMD Millipore). The clarified supernatant was concentrated at room temperature from a volume of 40 to 2 mL over approximately 4 h. The concentrated supernatant was flash frozen and stored at −20 °C prior to fractionation by analytical reversed phase high performance liquid chromatography (HPLC) using an Agilent 1260 Infinity equipped with a Phenomenex Luna 10 µm, C18(2) 100 Å, 250×4.6 mm column. The instrument was managed using Agilent Instrument 1 (Online) software. The mobile phase consisted of A: 0.1% trifluoroacetic acid (TFA, Acros Organics) in water and B: 0.1% TFA in 80% MeCN, 20% water and a gradient of 2% B from 0-5 min, 2-100% B from 5-50 min and 100% B from 50-60 min. A 0.5 µL aliquot of fractions eluting after 20 min was spotted with an equal volume of sinapic acid matrix and analyzed by MALDI-TOF-MS for the presence of venezuelin. Venezuelin-containing fractions (typically eluting at 35-38 min) were flash frozen and dried *in vacuo*. Three 1 mL injections of concentrated *S. katrae* ISP5550 supernatant yielded 100 µg of HPLC-purified, dry material.

3.4.5 Detection of (Me)Lan in venezuelin

HPLC-purified venezuelin was analyzed by GC/MS as previously described in Chapter 2. In brief, the sample was re-dissolved in 3 mL of 6 M HCl and heated at 110 °C for 10-12 h. The reaction was dried and mixed with a solution of acetyl chloride (Sigma-Aldrich) and methanol. Following reflux at 110 °C for 45 min, the reaction was dried and 3 mL of dichloromethane (Sigma-Aldrich) was added. After cooling the solution to 0 °C, 1 mL of pentafluoropropionic anhydride (Sigma-Aldrich) was added and the reaction was refluxed for 15 min at 110 °C. The reaction was dried and stored at −80 °C, and prior to GC/MS analysis the sample was re-dissolved in 100 µL of methanol. Particulates were removed by centrifugation at 23,700 ×g for 4 min. An Agilent Hewlett Packard 5973 mass spectrometer equipped with a CP-Chirasil-L Val, 25 m, 0.25 mm, 0.12 µm, 7 inch cage (Agilent Technologies) was used to perform the GC/MS analysis. The sample was introduced to the instrument via a split (1:5) injection of 2-5 µL with a

2.0 mL/min helium gas flow rate. The initial inlet temperature was 190 °C and a gradient of 20 °C/min from 160 °C for 5 min to 180 °C for 10 min was used. The following mass spectrometer settings were used: 185 °C for the MSD transfer line heater, 150 °C for the MS Quad analyzer, and 230 °C for the ion source. Data was acquired using scan and selected ion monitoring at 365 and 379 Da. The identity of (Me)Lan residues was confirmed by adding authentic (Me)Lan standards to the sample and analyzing the resulting mixture by GC/MS. The mixtures of sample and standard were prepared such that there was an approximately equal contribution of signal arising from the sample and standard.

3.5 References

1. Piper, C., Cotter, P.D., Ross, R.P., and Hill, C. (2009). Discovery of medically significant lantibiotics. *Curr. Drug Discov. Technol.* 6, 1-18.

2. Dischinger, J., Basi Chipalu, S., and Bierbaum, G. (2014). Lantibiotics: promising candidates for future applications in health care. *Int. J. Med. Microbiol.* 304, 51-62.

3. Cox, C.R., Coburn, P.S., and Gilmore, M.S. (2005). *Enterococcal* cytolysin: a novel two component peptide system that serves as a bacterial defense against eukaryotic and prokaryotic cells. *Curr. Protein Pept. Sci.* 6, 77-84.

4. Knerr, P.J., and van der Donk, W.A. (2012). Discovery, biosynthesis, and engineering of lantipeptides. *Annu. Rev. Biochem.* 81, 479-505.

5. Zhang, Q., Yu, Y., Velásquez, J.E., and van der Donk, W.A. (2012). Evolution of lanthipeptide synthetases. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18361-18366.

6. Zhang, Q., Yang, X., Wang, H., and van der Donk, W.A. (2014). High divergence of the precursor peptides in combinatorial lanthipeptide biosynthesis. *ACS Chem. Biol.* 9, 2686-2694.

7. Begley, M., Cotter, P.D., Hill, C., and Ross, R.P. (2009). Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins. *Appl. Environ. Microbiol.* 75, 5451-5460.

8. Marsh, A.J., O'Sullivan, O., Ross, R.P., Cotter, P.D., and Hill, C. (2010). *In silico* analysis highlights the frequency and diversity of type 1 lantibiotic gene clusters in genome sequenced bacteria. *BMC Genomics.* 11, 679-700.

9. Singh, M., and Sareen, D. (2014). Novel LanT associated lantibiotic clusters identified by genome database mining. *PLoS One.* 9, e91352.

10. Doroghazi, J.R., Albright, J.C., Goering, A.W., Ju, K.S., Haines, R.R., Tchalukov, K.A., Labeda, D.P., Kelleher, N.L., and Metcalf, W.W. (2014). A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol.* 10, 963-968.

11. Berdy, J. (2005). Bioactive microbial metabolites. *J. Antibiot.* 58, 1-26.

12. Ökesli, A., Cooper, L.E., Fogle, E.J., and van der Donk, W.A. (2011). Nine post-translational modifications during the biosynthesis of cinnamycin. *J. Am. Chem. Soc.* 133, 13753-13760.

13. Foulston, L.C., and Bibb, M.J. (2010). Microbisporicin gene cluster reveals unusual features of lantibiotic biosynthesis in actinomycetes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13461-13466.

14. Castiglione, F., Lazzarini, A., Carrano, L., Corti, E., Ciciliato, I., Gastaldo, L., Candiani, P., Losi, D., Marinelli, F., Selva, E., and Parenti, F. (2008). Determining the structure and mode of action of microbisporicin, a potent lantibiotic active against multiresistant pathogens. *Chem. Biol.* 15, 22-31.

15. Maffioli, S.I., Iorio, M., Sosio, M., Monciardini, P., Gaspari, E., and Donadio, S. (2014). Characterization of the congeners in the lantibiotic NAI-107 complex. *J. Nat. Prod.* 77, 79-84.

16. Iorio, M., Sasso, O., Maffioli, S.I., Bertorelli, R., Monciardini, P., Sosio, M., Bonezzi, F., Summa, M., Brunati, C., Bordoni, R., Corti, G., Tarozzo, G., Piomelli, D., Reggiani, A., and Donadio, S. (2014). A glycosylated, labionin-containing lanthipeptide with marked antinociceptive activity. *ACS Chem. Biol.* 9, 398-404.

17. Kodani, S., Hudson, M.E., Durrant, M.C., Buttner, M.J., Nodwell, J.R., and Willey, J.M. (2004). The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene *ramS* in *Streptomyces coelicolor*. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11448-11453.

18. Kersten, R.D., Yang, Y.L., Xu, Y., Cimermancic, P., Nam, S.J., Fenical, W., Fischbach, M.A., Moore, B.S., and Dorrestein, P.C. (2011). A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat. Chem. Biol.* 7, 794-802.

19. Zhang, Q., Doroghazi, J.R., Zhao, X., Walker, M.C., and van der Donk, W.A. (2015). Expanded natural product diversity revealed by analysis of lanthipeptide-like gene clusters in *Actinobacteria*. *Appl. Environ. Microbiol.* 81, 4339-4350.

20. Goto, Y., Li, B., Claesen, J., Shi, Y.X., Bibb, M.J., and van der Donk, W.A. (2010). Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. *PLoS Biol.* 8, e1000339.

21. Meindl, K., Schmiederer, T., Schneider, K., Reicke, A., Butz, D., Keller, S., Gühring, H., Vértesy, L., Wink, J., Hoffmann, H., Brönstrup, M., Sheldrick, G.M., and Süssmuth, R.D.

(2010). Labyrinthopeptins: a new class of carbacyclic lantibiotics. *Angew. Chem. Int. Ed.* 49, 1151-1154.

22. Völler, G.H., Krawczyk, J.M., Pesic, A., Krawczyk, B., Nachtigall, J., and Süssmuth, R.D. (2012). Characterization of new class III lantibiotics--erythreapeptin, avermipeptin and griseopeptin from *Saccharopolyspora erythraea*, *Streptomyces avermitilis* and *Streptomyces griseus* demonstrates stepwise N-terminal leader processing. *ChemBioChem.* 13, 1174-1183.

23. Krawczyk, J.M., Völler, G.H., Krawczyk, B., Kretz, J., Brönstrup, M., and Süssmuth, R.D. (2013). Heterologous expression and engineering studies of labyrinthopeptins, class III lantibiotics from *Actinomadura namibiensis*. *Chem. Biol.* 20, 111-122.

24. Krawczyk, B., Völler, G.H., Völler, J., Ensle, P., and Süssmuth, R.D. (2012). Curvopeptin: a new lanthionine-containing class III lantibiotic and its co-substrate promiscuous synthetase. *ChemBioChem.* 13, 2065-2071.

25. Küsters, E., Allgaier, H., Jung, G., and Bayer, E. (1984). Resolution of sulphur-containing amino acids by chiral phase gas chromatography. *Chromatographia.* 18, 287.

26. Ross, A.C., Liu, H., Pattabiraman, V.R., and Vederas, J.C. (2010). Synthesis of the lantibiotic lactocin S using peptide cyclizations on solid phase. *J. Am. Chem. Soc.* 132, 462-463.

27. Willey, J.M., and Gaskell, A.A. (2011). Morphogenetic signaling molecules of the streptomycetes. *Chem. Rev.* 111, 174-187.

28. Iftime, D., Jasyk, M., Kulik, A., Imhoff, J.F., Stegmann, E., Wohlleben, W., Süssmuth, R.D., and Weber, T. (2015). Streptocollin, a type IV lanthipeptide produced by *Streptomyces collinus* Tü 365. *ChemBioChem.* 16, 2615-2623.

Chapter 4. Characterization of a hybrid RiPP gene cluster from *Lachnospiraceae* C6A11[*]

4.1 Introduction

Bioinformatic surveys have revealed that LanM as well as YcaO proteins are encoded in a variety of genomic contexts (1, 2). Both enzyme classes are associated with RiPP biosynthesis, with LanM proteins catalyzing Ser/Thr dehydration and in most cases cyclization reactions with Cys residues (3, 4). YcaO proteins, synonymous to D proteins (5, 6), are associated with the installation of oxazol(ine) and thiazol(ine) post-translational modifications in a variety of substrates including linear azol(in)e-containing peptides, cyanobactins, thiopeptides, and bottromycins (7). Among the characterized examples of LanM and YcaO enzymes and systems, there are no reported cases of systems in which LanM and YcaO proteins coordinate their activities.

It was recently demonstrated that the coordination of oxazol(ine) and thiazol(ine) installing enzymes and lanthipeptide synthase proteins occurs during the biosynthesis of the thiopeptides (8). The representative Tbt system that generates the scaffold for thiomuracin consists of a set of proteins involved in dehydration (TbtB and C), two proteins involved in cyclodehydration (TbtF and G), one thiazoline dehydrogenase (TbtE), and a [4+2] cycloaddition enzyme (TbtD). TbtB and C bioinformatically correspond to the glutamylation and elimination domains of LanB proteins discussed in Chapter 1 and hence TbtB and C were designated as a split LanB. Through in vitro reconstitution, the order of post-translational modification was established as cyclodehydration of the substrate TbtA first by TbtF and G followed by dehydration by TbtB and C (8). In contrast to the dehydration reactions, cyclodehydration is a leader peptide dependent processes, while the former rely on a substrate conformation endowed by thiazole installation (9). Intriguingly, genome mining indicates the presence of LanM enzymes in clusters that also encode for YcaO-like proteins as shown in this chapter. Although it is tempting to project the findings from thiopeptide biosynthesis onto these systems, the fundamental variations between the two types of hybrid systems suggest possible differences in biosynthetic details.

[*] Studies were done in collaboration with Dr. Liujie Huo. Dr. Huo cloned the *lahA* genes and heterologously expressed them, purified the MBP-LahM1 fusion protein, and conducted the in vitro assays. X. Zhao cloned the genes for the LahM1, M2, and D1 proteins; heterologously co-expressed these proteins with various LahA peptides and optimized the production of the LahM2 protein.

4.2 Results

4.2.1 Bioinformatic identification and analysis of the Lah gene cluster

The *lah* gene cluster encoded within the genome of *Lachnospiraceae* C6A11 presents an intriguing compilation of RiPP post-translational modification genes (Figure 4.1A). The cluster can be independently identified from the results of BLAST queries of BsjM and FlvM1 or M2.

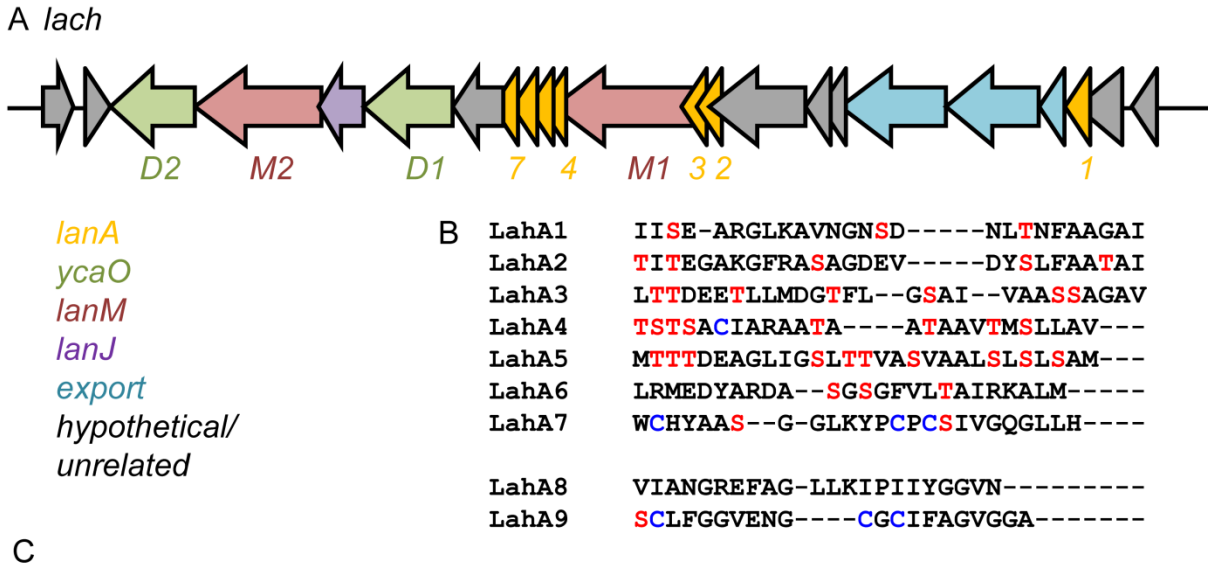

Figure 4.1. The *lah* gene cluster contains multiple enzymes and substrates. (A) *lah* gene cluster and (B) alignment of the LahA core peptides with Ser and Thr residues shown in red and Cys residues in blue. (C) LahA leader peptide alignment.

Although the pair of LanM proteins (annotated as LahM1 and M2) encoded within the *lah* gene cluster can be identified using BLAST, their domain architecture deviates from those of canonical LanM proteins (Figure 4.2). pFAM analysis of a prototypical LanM indicates the

presence of an approximately 400 amino acid long N-terminal domain of unknown function (Duf4135) and an approximately 350 amino acid long LanC-like domain. These annotated domains roughly correspond to the dehydratase (i.e. kinase and lyase) and cyclase domains within the CylM crystal structure that was discussed in Chapter 1. pFAM analysis of both LahM1 and M2 reveals the presence of a dehydratase domain. However, the dehydratase domain is shorter (approximately 200 amino acids in length) and a second (approximately 100 amino acid) dehydratase domain appears to be located immediately after (Figure 4.2A). Alignment of each of the LahM and canonical (CylM and LctM) dehydratase domains indicates conservation primarily in the region that corresponds to the N-lobe of the CylM crystal structure (10). Furthermore, the cyclase domain of both enzymes is not annotated by pFAM, and alignment with previously characterized LanM proteins confirms the absence of the Zn-binding residues conserved in prototypical LanM proteins (11). Inspection of the dehydratase domain in the alignment indicates conservation of residues characterized to be involved in phosphate elimination for LahM1 but not LahM2 (Figure 4.2B) (10, 12).

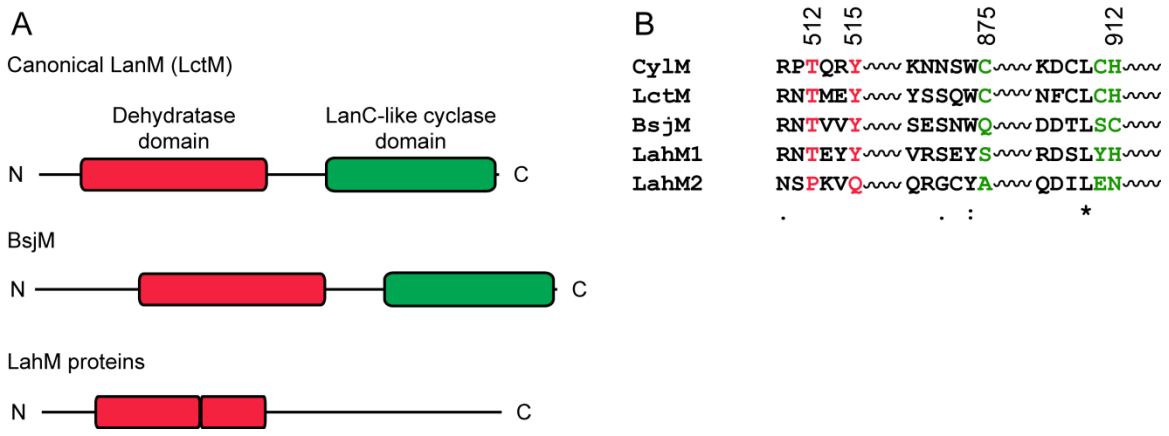

Figure 4.2. The LahM enzymes differ from canonical LanMs. (A) pFAM domain architecture of various LanM proteins compared to LahM1 and M2. (B) Alignment of canonical LanM enzymes (LctM and CylM), BsjM, and LahM1 and M2. Conserved residues that differ between the canonical enzymes and the others are shown in red for the dehydratase domain and green for the cyclization domain. Numbered residues correspond to the CylM sequence.

Additional enzymes encoded in the cluster include a flavin-dependent oxidoreductase ($LahJ_B$) and two YcaO proteins (Figure 4.1A). As mentioned in chapter 1, characterized LanJ orthologs are responsible for the reduction of dehydro amino acids to their corresponding D-amino acid counterparts (4, 13). The YcaO proteins, LahD1 and D2, following linear azol(in)e-containing peptides nomenclature (7), conspicuously lack cognate C proteins, which facilitate substrate recognition through the leader peptide (6, 14). Instead, LahD1 and D2 are reminiscent of the stand-alone YcaO proteins involved in bottromycin biosynthesis, and indeed bottromycin YcaO proteins are among the BLAST results with LahD1 and D2 as queries. Mutagenesis and metabolomics studies present evidence that the bottromycin YcaO proteins are responsible for cyclodehydrating and macrocyclizing the substrate (15). Unfortunately, absence of detailed in vitro characterization of these enzymes prevented assignment of the exact function of LahD1 and D2.

As in the case of the multi-substrate Proc and Flv systems discussed in Chapter 2, the putative LanA peptides (LahA1-7) are encoded within proximity of the LanM and YcaO encoding genes (Figure 4.1A). In addition, there are two additional putative LanA peptides encoded on a different scaffold in the draft genome and not within a clear RiPP biosynthetic context (LahA8 and 9), which can be uncovered from subjecting the leader peptide to BLAST. The leader peptides of all nine LahA peptides are conserved and bioinformatically annotated as Nif11 (Figure 4.1C). The conservation of the LahA leader peptides contrasts with the core peptides, which exhibit limited similarity and several distinctly lack Cys residues (Figure 4.1B). The general lack of Cys residues, and presence of only one Cys in LahA4 and three each in LahA7 and A9, is perhaps correlated with the Zn-ligand deficient cyclization domains of the LahM proteins (Figure 4.2B). Despite few Cys residues, the LahA core peptides contain a fair number of Ser and Thr residues (Figure 4.1B). Although it is not readily discernable which residues would be cyclodehydrated versus dehydrated.

Species from the genus *Lachnospiraceae* are anaerobic, Gram-positive isolates of disease-free mammalian digestive fauna (16-18). However, members of the genus are also implicated in the development of host disease states such as type 2 diabetes (19, 20) and resistance to *Clostridium difficile* colonization and thus infection (21). Both studies employed animal models, but did not present molecular deductions for the observed phenomena. Although a role of the Lah peptide products in mediating host disease or immunity is unclear, their

predicted modifications make them deserving candidates for biochemical characterization. Particularly intriguing aspects of the Lah system include: the division of biosynthetic labor coordinated by the suite of RiPP synthases, the biosynthetic order of post-translational modification events, the substrate tolerance implied by the multiple LahA peptides, the intra-substrate site specificity of the post-translational modification enzymes, and the final structures of the compounds. Investigation of these aspects required reconstitution of the activity the Lah enzymes and determination of substrate modification.

4.2.2 LahM proteins are active in the absence of other Lah enzymes

Initial exploration of the reconstitution of Lah enzyme activity focused on the arbitrarily chosen substrates LahA1 and A5 (Figure 4.1B and C). Co-expression of *lahA1* with either *lahM1* or *M2* resulted in a mass corresponding to unmodified peptide (Figure 4.3A). In contrast, co-expression of *lahA5* with *lahM1* resulted in three-fold dehydrated peptide and mono and di-phosphorylated peptide when co-expressed with *lahM2* (Figure 4.3B). Due to the observed modifications, subsequent efforts were directed to the LahA5 substrate.

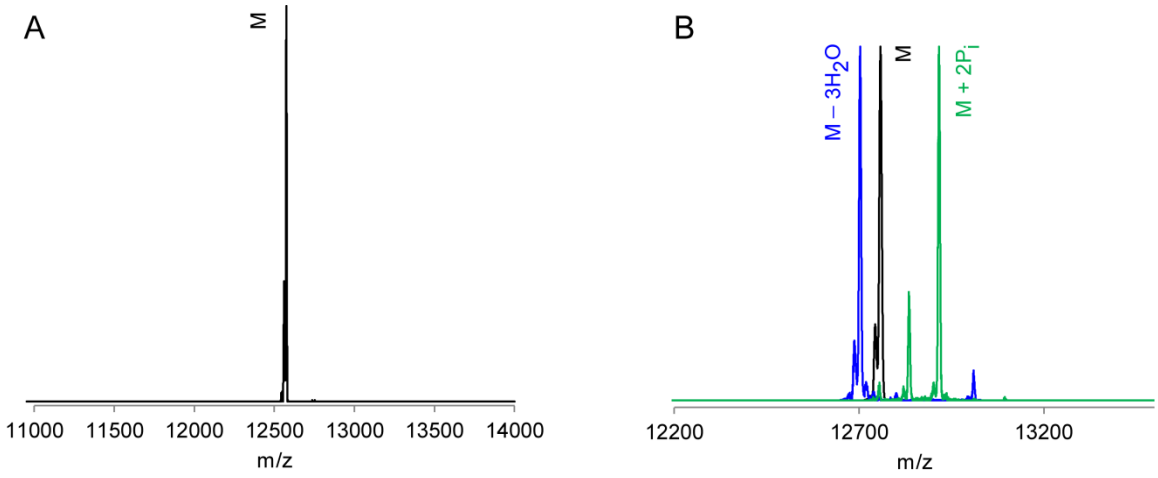

Figure 4.3. LahA5 is modified by both LahM1 and M2. (A) LahA1 co-expressed with either LahM1 or M2 is not modified, calculated average $[M+H]^+$ mass: 12573.6 m/z, observed average mass: 12574.3 m/z. (B) The spectrum for unmodified LahA5 is shown in black, calculated average $[M+H]^+$ mass: 12760.1 m/z, observed average mass: 12757.3 m/z. LahA5 co-expressed with LahM1, shown in blue, results in 3-fold dehydrated peptide, calculated average $[M+H]^+$ mass: 12706.1 m/z, observed average mass: 12702.9 m/z. LahA5 co-expressed with LahM2 is phosphorylated, diphosphorylated calculated average $[M+H]^+$ mass: 12920.1 m/z, observed average mass: 12915.7.

Next, co-expression studies of *lahA5* with binary combinations of the *lahM* or *lahD1* proteins did not yield additional modification beyond that observed in the co-expressions with just the LanM enzymes. The partial modification of LahA5 presented the possibility that the chosen substrate did not undergo cyclodehydration or that poor expression of the proteins was inhibiting activity. In order to investigate the latter possibility, production and purification of the LahM enzymes was undertaken.

The expression of the LahM proteins was first examined using His_6-tagged enzyme. Unfortunately, attempts to purify either His_6-LahM1 or M2 were hampered by low expression of the proteins, which was not remediated by the inclusion of rare tRNAs (22) or the molecular chaperones GroEL and ES (23) (Figure 4.4). However, expression and purification of a codon optimized, His_6-maltose-binding protein fusion of LahM2 (His_6-MBP-LahM2) was successful (24, 25). Correspondingly, His_6-MBP-LahM1 (without codon optimization) could also be purified. In vitro assays of LahA5 with either His_6-MBP-LahM1 or M2 resulted in the same modification extent as that observed in vivo.

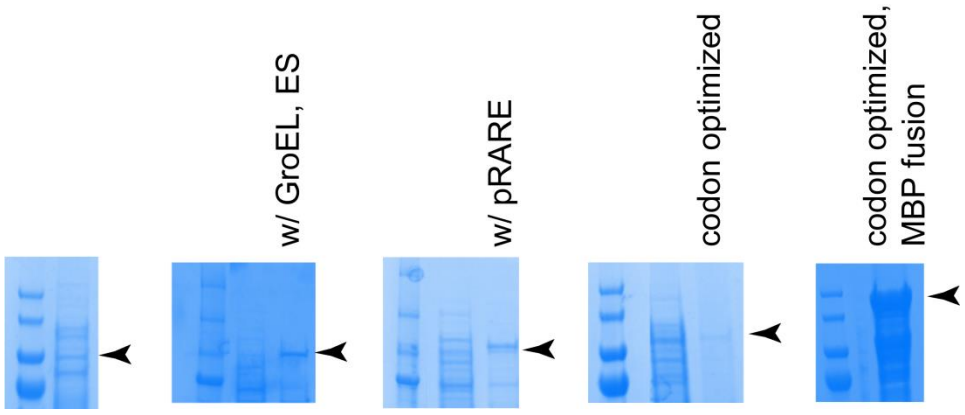

Figure 4.4. MBP improves LahM2 expression. The expected band size is indicated by the arrowhead.

4.3 Discussion

Several hypotheses can be presented regarding the biosynthetic order of the Lah system, first as in the case of the thiomuracin system, cyclodehydration must precede lanthipeptide synthase activity. Alternatively, the action of the lanthipeptide synthases occurs before the YcaO proteins or lastly the LahD and M enzymes function in tandem. The activity of the LahM

proteins observed for LahA5 is consistent with the latter two hypotheses (Figure 4.3B). Phosphorylation of LahA5 by M2 is perhaps not surprising given that the CylM Thr512 residue involved in phosphate elimination is a Pro390 in LahM2 and mutation of CylM Thr512 into Ala results in phosphorylated $CylL_S$ peptide (10), but it is unclear whether phosphorylation represents a modification in the LahA5 final product. However, the data do not exclude the possibility that the observed modification of LahA5 by LahM1 and M2 is incomplete and full modification requires a YcaO-generated intermediate of LahA5. The observed inactivity of LahD1 renders determination of its biosynthetic role and regioselectivity difficult. Future co-expression and in vitro studies employing both LahM1 and M2 will clarify the possibility of the enzymes functioning in tandem to affect modification. Due to the successful preparation of MBP-LahM2 and in vitro reconstitution, the purification of MBP fusions of the remaining Lah proteins is underway.

Given the similarity of the LahA leader peptides (Figure 4.1C) it is perhaps not surprising that the Lah proteins do not discern between discrete sets of substrates, as in the case of two-component lantibiotic systems (26). However, it seems impressive that each LahA substrate would require the activity of four to five enzymes for structural maturation. In addition, the six amino acid difference in length between the LahA1-5 and LahA6-9 peptides suggests leader peptide-mediated positioning of the substrate within the LahD or M active sites (27-29). These observations could argue against necessitating both LanM proteins for the modification of each substrate. Lastly, the absence of modification for LahA1 when co-expressed with either LahM1, M2 or D1 (Figure 4.3A) could be explained by LahM substrate discrimination being determined by preceding modification of the substrate by the LahD protein(s), which cannot be completely ruled out for this substrate.

4.4 Materials and methods

4.4.1 Culturing *Lachnospiraceae* C6A11 and gDNA isolation

Anaerobic culture of *L.* C6A11 was performed as described for *Ruminococcus flavefaceins* FD-1 in chapter 2, with modifications noted. The growth rate of *L.* C6A11 is slow, requiring ~7 d before reaching early lag phase. Thus, liquid and solid cultures were allowed to incubate for a minimum of 4 d at 37 °C before assessment by microscope. Furthermore, after liquid media was injected into agar slants, the slant was incubated for two-three d before transfer

to fresh liquid media. Initially, liquid RM02 and agar *L.* C6A11 slants were kindly provided by Dr. William Kelly's lab in AgResearch Limited in New Zealand and the liquid cultures were simply transferred to fresh media and allowed to incubate for 4-5 d. Upon establishment of an active *L.* C6A11 culture, gDNA was isolated using a MoBio kit, as per the manufacturer's instructions. *L.* C6A11 growth in the liquid media used to culture *R. flavefaciens* FD-1 in chapter 2 (approximately the same time scale as the R2YE media) was observed. Growth in agar slants of *R. flavefaciens* FD-1 media inoculated with *L.* C6A11 was also observed, thus agar slants of *L.* C6A11 were prepared using the same media for *R. flavefaciens* FD-1 agar slants described in chapter 2.

4.4.2 Cloning, expression, purification, and analysis of Lah peptides

See Chapter 2 for detailed descriptions, and any modifications are noted. Rather than using ligation reactions, Gibson assembly was used to assemble all plasmids (30). In preparation for Gibson assembly, backbone plasmid DNA was linearized using restriction endonucleases followed by treatment with CIP and purification using a QIAquick Gel Extraction Kit. Insert DNA obtained by PCR amplification was also purified using a QIAquick Gel Extraction Kit. In preparation for Gibson assembly, 5× Isothermal Reaction Buffer was prepared by dissolving 1.5 g PEG-8000 in 3 mL of 1 M Tris-HCl pH 7.5 with incubation in a 55 °C water bath, this solution was allowed to cool to room temperature before adding the remaining reagents (150 µL 2 M $MgCl_2$, 60 µL of 100 mM dGTP, 60 µL of 100 mM dCTP, 60 µL of 100 mM dATP, 60 µL of 100 mM dTTP, 300 µL of 1M DTT, 300 µL of 100 mM NAD^+, and water to a final volume of 6 mL). Then the Gibson assembly reaction mixture was prepared by combining 320 µL of 5× Isothermal Reaction Buffer, 860 µL of sterile water, 0.64 µL of T5 Exonuclease, and 20 µL of Phusion Polymerase. Then, 15 µL of the Gibson assembly reaction mixture was combined with 50-100 ng of backbone DNA and a 30-50 fold excess (normalized to ng/kb) of insert DNA. The Gibson assembly reactions were incubated at 50 °C for 1 h and then used to transformed chemically competent *E. coli* DH10B (or plasmid maintenance cell line). Since a variety of plasmid backbones were used, the final backbone is noted in the primer table for each set of primers. A table of the prepared constructs and a brief description is also provided.

Table 4.1. Primers for cloning *lah* genes.

Primer sequence 5′ to 3′	Annealing temperature, °C	Primer direction, construct name_plasmid backbone
gtataagaaggagatatacatATGAAGCAGAATGAATTCTACGATTATATATC	60	Forward, LahM1_mcs2_pACYC
gcagcggtttctttaccagactcgagTCATGTTAATTCCTCCTTAGTTTTTTTTATTG	60	Reverse, LahM1_mcs2_pACYC
gttaagtataagaaggagatatacatATGAACTTAACACTGAAGCAATGGCTGG	60	Forward, LahM2_mcs2_pACYC
gcagcagcggtttctttaccagactcgagCTACTTCGACAGCCATCCCAAAAATAAAAC	60	Reverse, LahM2_mcs2_pACYC
gtatattagttaagtataagaaggagatatacatatgAAAATCGAAGAAGGTAAACTG	62	Forward, MBP-LahYcao1_mcs2_pCDF1
GGTGTTCTGTCCTTATATTTTGAATAATTACTCTCGAGGGATCCGGATTG	62	Reverse, MBP-LahYcao1_mcs2_pCDF1
CAATCCGGATCCCTCGAGAGTAATTATTCAAAATATAAGGACAGAACACC	62	Forward, MBP-LahYcao1_mcs2_pCDF1
gcagcggtttctttaccagactcgagTTAATATTCTCTGCCGTATATTTCCTTTAATTC	62	Reverse, MBP-LahYcao1_mcs2_pCDF1
cctggtgccgcgcggcagccatatgAAGCAGAATGAATTCTACGATTATATATCCG	67	Forward, His_6-LahM1_pET15b
gctttgttagcagccggatccTCATGTTAATTCCTCCTTAGTTTTTTTTATTGC	67	Reverse, His_6-LahM1_pET15b
gcagcggcctggtgccgcgcggcagccatatgAACTTAACACTGAAGCAATGGCTGGAG	67	Forward, His_6-LahM2_pET15b
cctttcgggctttgttagcagccggatccCTACTTCGACAGCCATCCCAAAAATAAAAC	67	Reverse, His_6-LahM2_pET15b
catcacagcagcggcctggtgccgcgcggcagccatAAAATCGAAGAAGGTAAACTG	62	Forward, His_6-MBP-LahM2_codon optimized_pET15b
CTTCAAACAGTTCCAGCCATTGCTTCAGGGTCAGGTTAGTCTGCGCGTCTTTCAGG	62	Reverse, His_6-MBP-LahM2_codon optimized_pET15b
CAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTAACCTGACCCTGAAGCAATGG	62	Forward, His_6-MBP-LahM2_codon optimized_pET15b

Table 4.1. (cont.)

ctcagcttcctttcgggctttgttagcagccggatccTTACTTGCTCA GCCAGCCC	62	Reverse, His_6-MBP-LahM2_codon optimized_pET15b

Primers for the substrates were designed by Dr. Huo.

Table 4.2. Application of selected Lah constructs.

Construct	Comments
His_6-LahA_mcs1_pRSF	For (co-)expression and purification
LahM1_mcs2_pACYC	For co-expression with *lahA* genes
LahM2_mcs2_pACYC	For co-expression with *lahA* genes
MBP-LahYcao1_mcs2_pCDF1	For co-expression with *lahA* genes
His_6-LahM1_pET15b	For purification and in vitro studies
His_6-LahM2_pET15b	For purification and in vitro studies
His_6-MBP-LahM2_codon optimized_pET15b	For purification and in vitro studies

Small scale pilot expressions of the *lahA* peptide genes were carried out prior to larger expressions. In these cases, a single colony transformant was inoculated into 30 mL of liquid LB culture containing the appropriate antibiotic(s). The culture was incubated with shaking at 37 °C until an OD_{600} of 0.6-0.8 was achieved at which point IPTG was added to a final concentration of 50 µM. The culture was then incubated with shaking at 18 °C for 18-20 h and then harvested and stored until lysis. Larger expressions were carried out as described in Chapter 2 but with a post-induction temperature of 18 °C rather than 37 °C since primarily unmodified LahA5 peptide was observed from co-expressions employing a 37 °C post-induction temperature.

4.4.3 Expression and purification of Lah proteins

Expressions were carried out as described in Chapter 2 but with a post-induction temperature of 18 °C rather than 37 °C. Protein purifications were carried out as previously described for LanM proteins (31) with the following modifications. The cell pellet was lysed using an Avestin EmulsiFlex-C3, and an Äkta FPLC was used to elute the protein during IMAC.

4.5 References

1. Cox, C.L., Doroghazi, J.R., and Mitchell, D.A. (2015). The genomic landscape of ribosomal peptides containing thiazole and oxazole heterocycles. *BMC Genomics.* 16, 778-794.

2. Zhang, Q., Yu, Y., Velásquez, J.E., and van der Donk, W.A. (2012). Evolution of lanthipeptide synthetases. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18361-18366.

3. Knerr, P.J., and van der Donk, W.A. (2012). Discovery, biosynthesis, and engineering of lantipeptides. *Annu. Rev. Biochem.* 81, 479-505.

4. Yang, X., and van der Donk, W.A. (2015). Post-translational introduction of D-alanine into ribosomally synthesized peptides by the dehydroalanine reductase NpnJ. *J. Am. Chem. Soc.* 137, 12426-12429.

5. Dunbar, K.L., and Mitchell, D.A. (2013). Revealing nature's synthetic potential through the study of ribosomal natural product biosynthesis. *ACS Chem. Biol.* 8, 1083-1083.

6. Dunbar, K.L., Melby, J.O., and Mitchell, D.A. (2012). YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. *Nat. Chem. Biol.* 8, 569-575.

7. Arnison, P.G., Bibb, M.J., Bierbaum, G., Bowers, A.A., Bugni, T.S., Bulaj, G., Camarero, J.A., Campopiano, D.J., Challis, G.L., Clardy, J., Cotter, P.D., Craik, D.J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P.C., Entian, K.D., Fischbach, M.A., Garavelli, J.S., Göransson, U., Gruber, C.W., Haft, D.H., Hemscheidt, T.K., Hertweck, C., Hill, C., Horswill, A.R., Jaspars, M., Kelly, W.L., Klinman, J.P., Kuipers, O.P., Link, A.J., Liu, W., Marahiel, M.A., Mitchell, D.A., Moll, G.N., Moore, B.S., Müller, R., Nair, S.K., Nes, I.F., Norris, G.E., Olivera, B.M., Onaka, H., Patchett, M.L., Piel, J., Reaney, M.J.T., Rebuffat, S., Ross, R.P., Sahl, H.G., Schmidt, E.W., Selsted, M.E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Süssmuth, R.D., Tagg, J.R., Tang, G.L., Truman, A.W., Vederas, J.C., Walsh, C.T., Walton, J.D., Wenzel, S.C., Willey, J.M., and van der Donk, W.A. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* 30, 108-160.

8. Hudson, G.A., Zhang, Z., Tietz, J.I., Mitchell, D.A., and van der Donk, W.A. (2015). In vitro biosynthesis of the core scaffold of the thiopeptide thiomuracin. *J. Am. Chem. Soc.* 137, 16012-16015.

9. Zhang, Z., Hudson, G.A., Mahanta, N., Tietz, J.I., van der Donk, W.A., and Mitchell, D.A. (2016). Biosynthetic timing and substrate specificity for the thiopeptide thiomuracin. *J. Am. Chem. Soc.* 138, 15511-15514.

10. Dong, S.H., Tang, W., Lukk, T., Yu, Y., Nair, S.K., and van der Donk, W.A. (2015). The enterococcal cytolysin synthetase has an unanticipated lipid kinase fold. *elife.* 4, e07607.

11. Paul, M., Patton, G.C., and van der Donk, W.A. (2007). Mutants of the zinc ligands of lacticin 481 synthetase retain dehydration activity but have impaired cyclization activity. *Biochemistry*. 46, 6268-6276.

12. You, Y.O., and van der Donk, W.A. (2007). Mechanistic investigations of the dehydration reaction of lacticin 481 synthetase using site-directed mutagenesis. *Biochemistry.* 46, 5991-6000.

13. Huo, L., and van der Donk, W.A. (2016). Discovery and characterization of bicereucin, an unusual D-amino acid-containing mixed two-component lantibiotic. *J. Am. Chem. Soc.* 138, 5254-5257.

14. Burkhart, B.J., Hudson, G.A., Dunbar, K.L., and Mitchell, D.A. (2015). A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nat. Chem. Biol.* 11, 564-570.

15. Crone, W.J.K., Vior, N.M., Santos-Aberturas, J., Schmitz, L.G., Leeper, F.J., and Truman, A.W. (2016). Dissecting bottromycin biosynthesis using comparative untargeted metabolomics. *Angew. Chem., Int. Ed.* 55, 9638-9642.

16. McLellan, S.L., Newton, R.J., Vandewalle, J.L., Shanks, O.C., Huse, S.M., Eren, A.M., and Sogin, M.L. (2013). Sewage reflects the distribution of human faecal *Lachnospiraceae*. *Environ. Microbiol.* 15, 2213-2227.

17. Sagheddu, V., Patrone, V., Miragoli, F., Puglisi, E., and Morelli, L. (2016). Infant early gut colonization by *Lachnospiraceae*: high frequency of *Ruminococcus gnavus*. *Front. Pediatr.* 4, 10.3389/fped.2016.00057.

18. Meehan, C.J., and Beiko, R.G. (2014). A phylogenomic view of ecological specialization in the Lachnospiraceae, a family of digestive tract-associated bacteria. *Genome Biol. Evol.* 6, 703-713.

19. Zeng, H., Ishaq, S.L., Zhao, F.Q., and Wright, A.D. (2016). Colonic inflammation accompanies an increase of beta-catenin signaling and *Lachnospiraceae*/*Streptococcaceae* bacteria in the hind gut of high-fat diet-fed mice. *J. Nutr. Biochem.* 35, 30-36.

20. Kameyama, K., and Itoh, K. (2014). Intestinal colonization by a *Lachnospiraceae* bacterium contributes to the development of diabetes in obese mice. *Microbes Environ.* 29, 427-430.

21. Reeves, A.E., Koenigsknecht, M.J., Bergin, I.L., and Young, V.B. (2012). Suppression of *Clostridium difficile* in the gastrointestinal tracts of germfree mice inoculated with a murine isolate from the family *Lachnospiraceae*. *Infect. Immun.* 80, 3786-3794.

22. Novy, R., Drott, D., Yaeger, K., and Mierendorf, R. (2001). Overcoming the codon bias of *E. coli* for enhanced protein expression. *Newsletter of Novagen, Inc.*

23. Thomas, J.G., Ayling, A., and Baneyx, F. (1997). Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from *E. coli*. To fold or to refold. *Appl. Biochem. Biotechnol.* 66, 197-238.

24. Diguan, C., Li, P., Riggs, P.D., and Inouye, H. (1988). Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene.* 67, 21-30.

25. Kapust, R.B., and Waugh, D.S. (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* 8, 1668-1674.

26. Thibodeaux, G.N., McClerren, A.L., Ma, Y., Gancayco, M.R., and van der Donk, W.A. (2015). Synergistic binding of the leader and core peptides by the lantibiotic synthetase HalM2. *ACS Chem. Biol.* 10, 970-977.

27. Thibodeaux, C.J., Wagoner, J., Yu, Y., and van der Donk, W.A. (2016). Leader peptide establishes dehydration order, promotes efficiency, and ensures fidelity during lacticin 481 biosynthesis. *J. Am. Chem. Soc.* 138, 6436-6444.

28. Koehnke, J., Mann, G., Bent, A.F., Ludewig, H., Shirran, S., Botting, C., Lebl, T., Houssen, W.E., Jaspars, M., and Naismith, J.H. (2015). Structural analysis of leader peptide binding enables leader-free cyanobactin processing. *Nat. Chem. Biol.* 11, 558-563.

29. Czekster, C.M., Ge, Y., and Naismith, J.H. (2016). Mechanisms of cyanobactin biosynthesis. *Curr. Opin. Chem. Biol.* 35, 80-88.

30. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods.* 6, 343-341.

31. Li, B., Cooper, L.E., and van der Donk, W.A. (2009). Chapter 21. In vitro studies of lantibiotic biosynthesis. *Methods Enzymol.* 458, 533-558.

Chapter 5. Investigations into the sLanB-encoding gene cluster of *Bacillus halodurans* C-125[*]

5.1 Introduction

In vitro reconstitution of the full length LanB enzyme, NisB, which is involved in the biosynthesis of the lantibiotic nisin, provided unprecedented insight into the mechanism of LanB enzymes (1). NisB was demonstrated to add a glutamyl group, derived from aminoacylated-$tRNA^{Glu}$, to the Ser and Thr residues of the NisA precursor peptide and eliminate the glutamate to form Dha and Dhb. This mechanism diverges from that of the class II-IV synthetases discussed in Chapter 1. The insights gained from NisB provide a backdrop for understanding LanB-like enzymes encoded in non-lanthipeptide gene cluster contexts. Indeed, in vitro reconstitution and confirmed dependence on aminoacyl-tRNA (aa-tRNA) was forthcoming for the canonical class I synthase lantibiotic synthase MibB (2) and the split LanB enzymes TbtB and C (discussed in further detail in Chapter 4) (3). Perhaps even more intriguing are the instances of LanB-like enzymes that lack a discernable elimination domain, which in NisB eliminates glutamate from the glutamylated intermediate to produce dehydrated substrate. Although bioinformatics was critical in identifying these sLanB enzymes (1), additional mechanistic and structural ramifications could not be deduced in the absence of more detailed investigations. Thus, a sLanB-containing gene cluster was identified and heterologous expression of the cluster as well as mechanistic investigation of one of the sLanB enzymes was attempted.

5.2 Results

5.2.1 The Hdn sLanB proteins bear hallmarks of LanB proteins

The proposed *hdn* gene cluster in *B. halodurans* C-125 consists of 18 genes, seven of which encode for sLanB enzymes (Figure 5.1A). Although an N-terminal lanthipeptide dehydratase domain is annotated in some of the HdnB proteins, pFAM analysis also reveals a conspicuous absence of a C-terminal lanthipeptide dehydratase domain for all seven enzymes (Figure 5.1B). The annotated N-terminal lanthipeptide dehydratase domain corresponds to the

[*] Studies were done in collaboration with Dr. Rebecca Splain. Dr. Splain cloned the *hdn* genes, heterologously expressed and purified them, and conducted the in vitro assays. X. Zhao cloned portions of the *hdn* gene cluster and heterologously expressed the cloned gene cluster. X. Zhao also prepared the materials for and performed the phosphorus NMR studies.

glutamylation domain in the NisB crystal structure, which esterifies the Ser and Thr residues in the NisA precursor with glutamate from glutamyl-tRNA^{Glu} (1). Subsequently, the NisB elimination domain catalyzes the elimination of glutamate from NisA to yield dehydro amino acids.

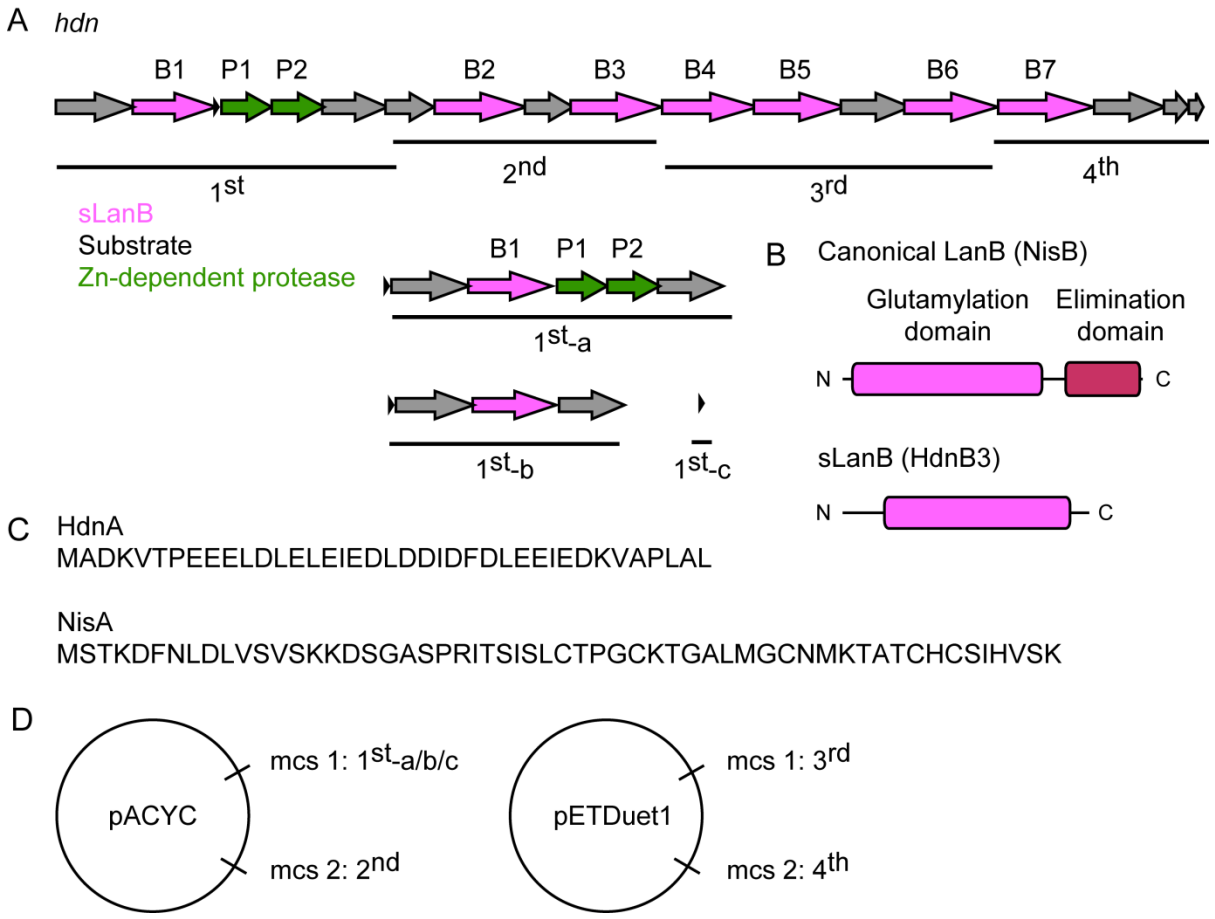

Figure 5.1. The *hdn* gene cluster encodes multiple sLanB proteins. (A) *hdn* gene cluster, with the HdnB proteins colored magenta and proteases colored green. The lines below the cluster denote portions that were cloned together for co-expression studies. The first portion of the cluster was altered such that a *his*$_6$*-hdnA* gene was placed before the other genes in the *hdn* cluster while the endogenous *hdnA* was abolished (inset) (B) Domain architecture comparison of full length and sLanB proteins, showing the lack of elimination domain in HdnB3. (C) Amino acid sequence of HdnA and NisA. (D) Maps of the constructs used for the co-expression studies (also see Figure 5.3), primer sequences can be found in Section 5.4. Complementary constructs in which the first or second mcs was empty were also prepared.

Despite the presence of several sLanB enzymes in the *hdn* gene cluster, only a single gene encoding for a 39 amino acid peptide (HdnA) was detected in the cluster. Intriguingly, the residue composition of HdnA differs from that of prototypical LanB precursor peptides (Figure 5.1C). Specifically, compared to NisA, HdnA is shorter in length, lacks any Ser, Thr, and Cys residues, and exhibits an abundance of Asp and Glu residues (17 out of 39 amino acids). In addition to the Hdn sLanB proteins, two Zn-dependent proteases (HdnP1 and 2) are annotated (Figure 5.1A), but their exact roles in biosynthesis were unclear. The profusion of and mechanistic implications of the sLanB enzymes in the *hdn* gene cluster presented an intriguing opportunity to extend the understanding of LanB-like enzymes. Thus, investigation of the enzymes and the resultant product of the *hdn* gene cluster was undertaken.

5.2.2 sLanB enzymes add amino acids to the C-terminus of a substrate peptide

All of the Hdn sLanB enzymes (1-7) were expressed as soluble His_6 fusion proteins. In vitro assays of combining HdnA with *E. coli* cell extract, ATP, Mg^{2+}, glutamate, and TCEP resulted in an increase in mass of the HdnA precursor peptide (Figure 5.2A). Subsequent in vitro assays with deuterated amino acids suggested that HdnB1 generated HdnA+Ala and HdnB7 accepted this as a substrate to produce HdnA+AlaTrp (Figure 5.2B-D). These observations prompted investigation of whether the HdnB enzymes functioned within a heterologous host.

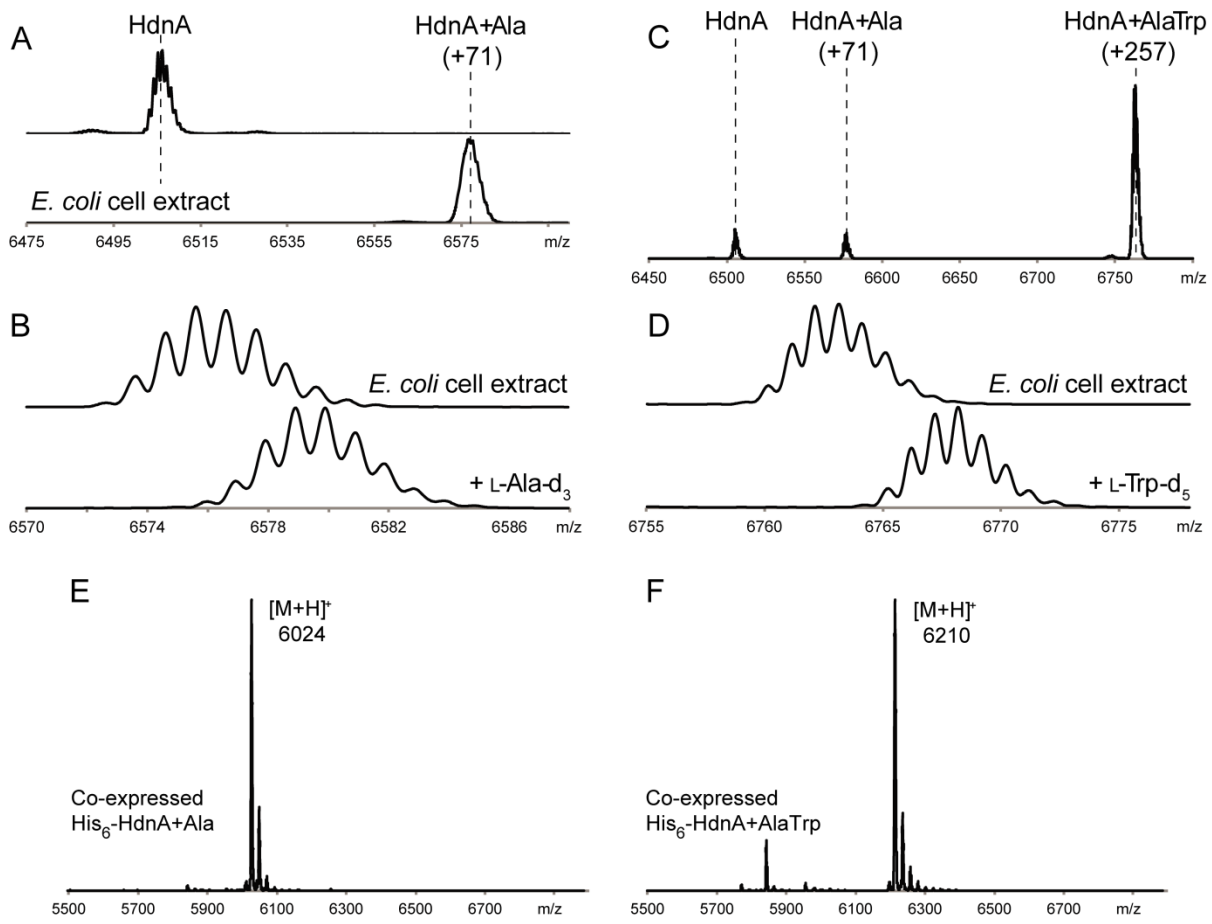

Figure 5.2. Addition of Ala and Trp to HdnA. In vitro assays using cell extract of (A) HdnA and B1, (B) HdnA+Ala and HdnB7, and (C, D) deuterated amino acids. Co-expressions of (E) His_6-HdnA and B1 and (F) His_6-HdnA, B1, and B7.

5.2.3 Co-expression of the *hdn* genes

Co-expression of *hdnA* and *B1* yielded a peptide with a mass corresponding to the HdnA+Ala (Figure 5.2E). In addition, co-expressing the genes encoding for HdnA, B1 and B7, yielded a peptide mass corresponding to HdnA+AlaTrp (Figure 5.2F). Given the successful reconstitution of Hdn protein activity in *E. coli*, co-expressions encompassing additional portions of the *hdn* gene cluster were attempted (Figure 5.3 and see Figure 5.2A for gene cluster portions). However, co-expressing additional portions of the *hdn* gene cluster did not result in additional mass changes (Figure 5.3).

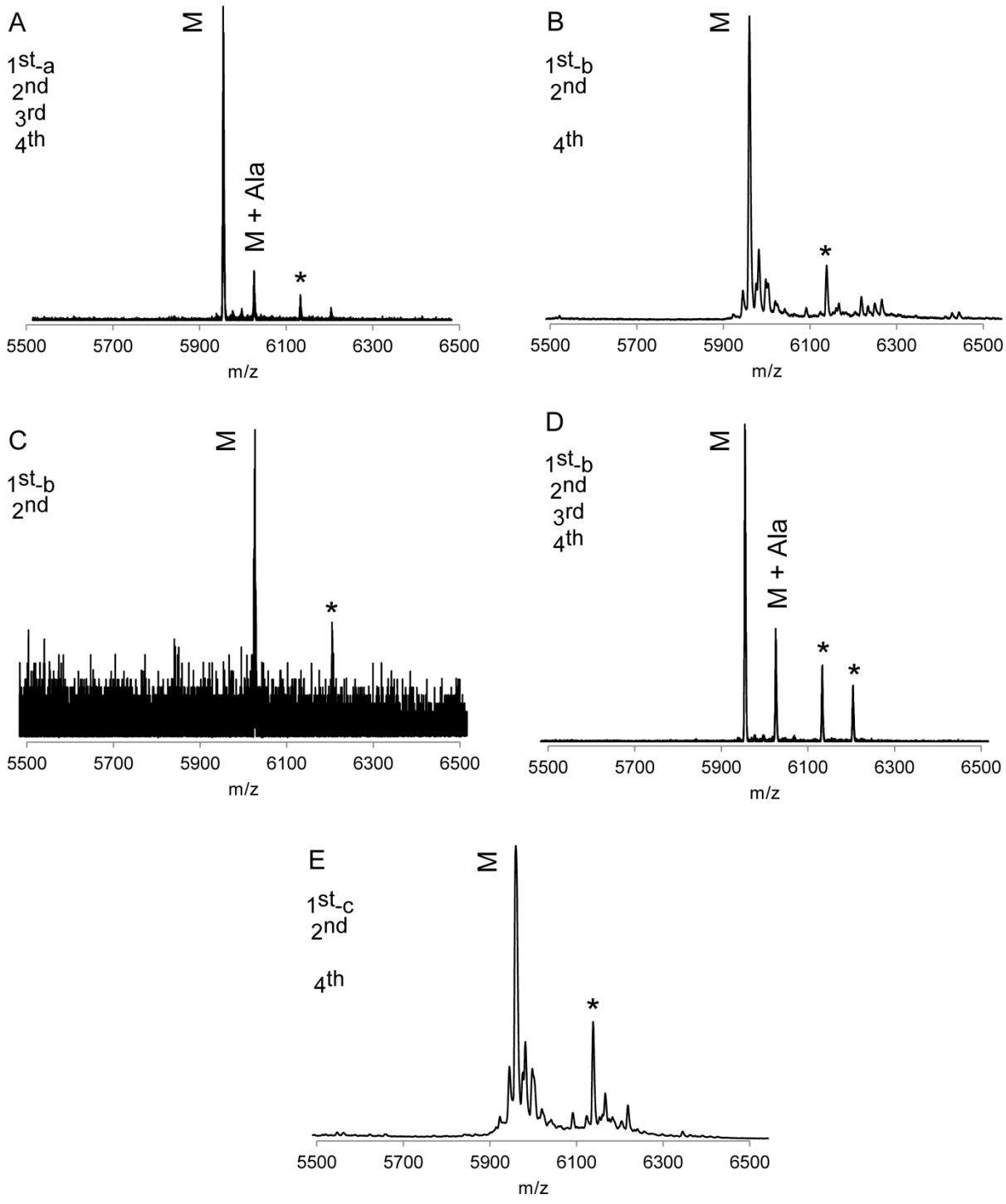

Figure 5.3. Co-expression of portions of the *hdn* gene cluster. Peaks with masses corresponding to unmodified HdnA peptide are labeled "M". Portions of the gene cluster that were co-expressed are shown in detail in Figure 5.1. Co-expression of (A) the full *hdn* gene cluster, (B) the first, second, and fourth portions of *hdn* gene cluster without *hdnP1* and *P2*, (C) the first two portions of *hdn* gene cluster without *hdnP1* and *P2*, (D) the *hdn* gene cluster without

Figure 5.3. (cont.)

hdnP1 and *P2*, and (E) second and fourth portions of *hdn* gene cluster with *hdnA*. * denotes a mass corresponding to gluconoylated peptide. Spectra were acquired from samples that were subjected to IMAC.

5.2.4 Mechanistic investigations of HdnB7

Inspired by the mechanistic investigations of other RiPP enzymes, analogous studies were undertaken for HdnB7 (4, 5). First, the dependence of HdnB1 on ATP was evaluated by supplying aa-tRNA in the absence of ATP. However, addition of Ala to HdnA was not observed until ATP was also included. This suggested that ATP was used to activate the carboxylate of HdnA, and it was envisioned that incorporation of isotopically labelled Ala into the C-terminus of the HdnA peptide would provide a handle that could inform on the mechanism of catalysis.

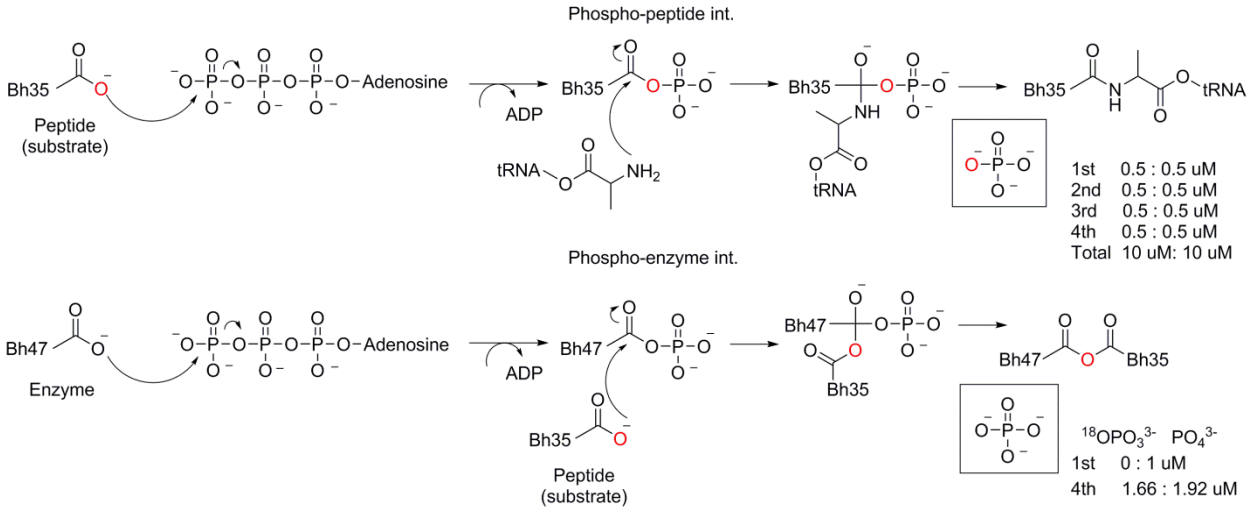

Figure 5.4. Two proposed mechanisms of catalysis for HdnB7, the C-terminal ^{18}O isotope label in HdnA (incorporated using HdnB1) is colored red.

Ala bearing doubly labelled oxygen-18 (Ala-$^{18}O_2$) at its carboxylate was selected for its minimal perturbation to the substrate, and phosphorus-31 nuclear magnetic resonance spectroscopy (^{31}P NMR) was selected by precedent (6). Two mechanisms (Figure 5.4) were proposed, operation of the first mechanism, which proceeds through a phosphopeptide intermediate, would result in the production of a 50:50 mixture of $^{18}OPO_3^{3-}$ and PO_4^{3-} under both single- and multiple turnover

conditions. In contrast the second proposed mechanism would yield exclusively $^{18}OPO_3^{3-}$ under single-turnover conditions (limiting peptide) and a mixture of $^{18}OPO_3^{3-}$ and PO_4^{3-} under multiple turnover conditions. Although a single phosphate peak was observed under single-turnover conditions, a single phosphate peak was also observed under multiple turnover conditions of HdnB7, rendering substantiation of either mechanism difficult.

5.3 Discussion

Despite an increasing appreciation for the ubiquity of aa-tRNA-dependent enzymes involved in secondary metabolism (7, 8), few share annotated sequence similarities with the HdnB enzymes. However, an analogous biosynthetic gene cluster was highlighted in recent work by the Moore lab on the biosynthesis of pyrroloquinoline alkaloids from the marine bacterium *Streptomyces sp.* CNR-698 (9). Production of the small molecule ammosamide was abolished upon deleting genes within the *amm* gene cluster, which shares features with the *hdn* cluster. Specifically, the *amm* gene cluster from *S. sp.* CNR-698 features four annotated sLanB proteins as well as an acidic precursor peptide (encoded by *amm6*) that exhibits conservation with the C-terminus of HdnA. Unfortunately, the in vitro activity of the Amm sLanB proteins was not reconstituted, and thus their role in ammosamide biosynthesis remains unclear. Based on the findings of the Moore lab, it is tempting to speculate that a similar underlying biosynthetic logic is operative in the Hdn system, in which a ribosomally-encoded peptide is the biosynthetic origin of a small molecule final product. However, it cannot be excluded that the precursor peptide of the Amm system is tailored into a final product that regulates the production of ammosamides.

The observation of primarily unmodified HdnA upon expression of portions of the Hdn gene cluster is puzzling (Figure 5.3). However, this observation could be consistent with a model in which HdnA serves as a scaffold that is modified, proteolyzed, and regenerated (possibly by the sLanB proteins). Alternatively, lack of further HdnA peptide modification could be the result of unsuccessful enzyme expression in *E. coli*. Indeed, attempts to purify either His_6-HdnP protease resulted in insoluble protein. The current heterologous host may be a less than ideal platform for the expression of the *hdn* gene cluster, thus hampering full pathway reconstitution. Alternative approaches to circumvent host incompatibility could include genetic manipulation of *B. halodurans* C-125 or introduction of the Hdn gene cluster into a *Bacillus* heterologous host. The primacy of the former approach is evidenced by the studies on the ammosamides.

The lack of discrete $^{18}OPO_3^{3-}$ and PO_4^{3-} peaks in the ^{31}P NMR spectrum of the HdnB7 reaction can be attributed to insufficient resolution or production of phosphate that is not coupled to substrate turnover. The latter possibility is more likely since control studies in which $^{18}OPO_3^{3-}$ and PO_4^{3-} was generated and prepared by 17x-PTDH and added to the HdnB7 in vitro assay sample buffer (see Section 5.4 for more details) yielded clearly distinct $^{18}OPO_3^{3-}$ and PO_4^{3-} peaks. Furthermore, the possibility of $^{18}OPO_3^{3-}$ signal suppression by PO_4^{3-} contamination was excluded by monitoring for the potential presence of PO_4^{3-} in the individual components of the HdnB7 in vitro reaction. No PO_4^{3-} contamination was observed if the tRNA was purified by denaturing, acidic polyacrylamide gel as previously described for aa-tRNA (10). However, the production of phosphate by HdnB7 was not monitored using ^{31}P NMR in the absence of full length HdnA peptide nor in the presence of HdnA leader peptide. It is anticipated that in both of these cases the production of a nontrivial amount of phosphate would be observed. Future attempts to examine the mechanistic details of sLanB enzymes should be preceded by evaluation of ATP-hydrolysis activity under various conditions, prior to generation of isotopically labelled substrate. Phosphate production should be examined in absence of both the peptide substrate and aa-tRNA, the presence of peptide substrate and absence of aa-tRNA, and vice versa.

5.4 Materials and methods

5.4.1 General materials and methods

Reagents were the same as those specified in Chapter 2. Sodium phosphate monobasic, monohydrate was purchased from Fischer Scientific and adenosine di- and triphosphate hydrates were purchased from Sigma-Aldrich.

5.4.2 Culturing *B. halodurans* C-125 and gDNA isolation

All procedures were performed aseptically. *B. halodurans* was maintained as a 40% glycerol stock at −80 °C. The glycerol stock was revived by first streaking onto solid LB media and incubating at 37 °C for 16-20 h. A single colony of the organism was transferred into 5 mL of liquid LB media and incubated at 37 °C, 200-220 rpm for ~24 h. gDNA was isolated using a MoBio kit, as per the manufacturer's instructions.

5.4.3 Cloning, expression, and analysis of the *hdn* gene cluster

See Chapter 2 for detailed descriptions, and any modifications are noted in Chapter 4. Specifically, Gibson Assembly was used to assemble all plasmids (11). Single restriction endonuclease digestion of the *hdn* gene cluster constructs indicated incorporation of inserts of the correct size, and sequencing of the 5' and 3' extremities of each mcs also indicated the expected insert sequences. However, full sequencing coverage of each portion of the *hdn* cluster was not achieved (a typical sequencing run provided ~1 kb of data), and full sequencing coverage is necessary to rule out the unlikely possibility of point mutations in the constructs. The HdnA-bearing construct was designed such that a *his*$_6$*-hdnA* gene was placed before the remaining genes in the *hdn* cluster and the endogenous copy of *hdnA* was removed (inset).

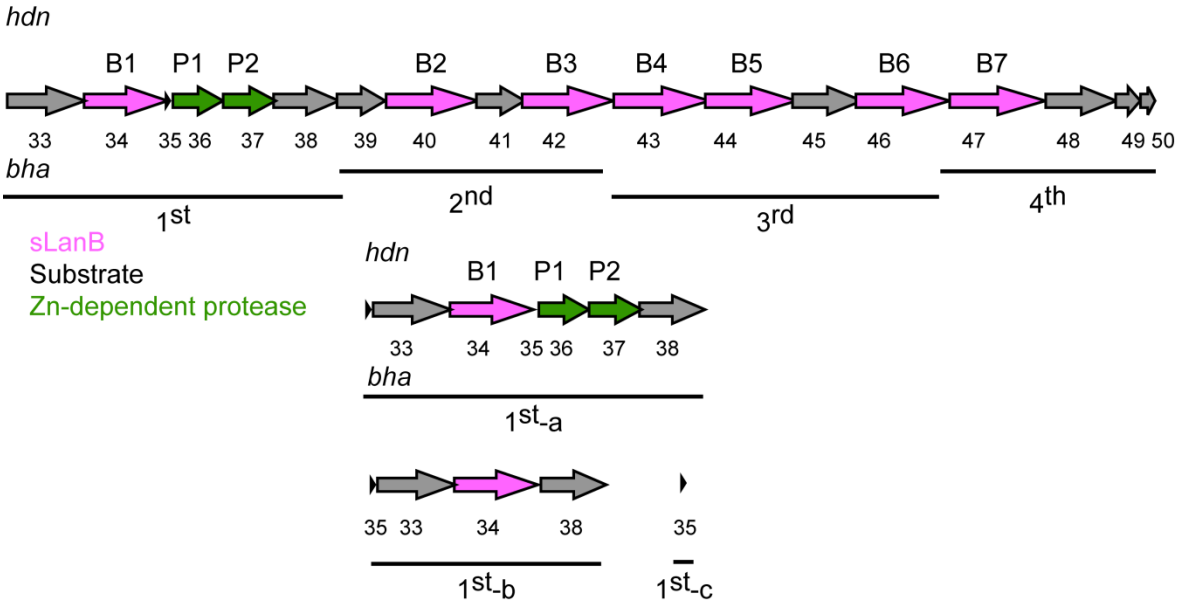

Figure 5.5. *hdn* gene cluster key for cloning.

Table 5.1. Primers for cloning *hdn* genes.

Primer sequence 5′ to 3′	Annealing temperature, °C	Primer direction, construct name_ plasmid backbone_ restriction site/ overlap sequence
gcagcagccatcaccatcatcaccacagccaggatccaATGGCCGAC AAGGTTACACC	66	Forward, HisBh35_mcs1_pACYC_ BamHI

Table 5.1. (cont.)

ggcgcgccgagctcgaattcTTAAAGAGCCAGAGGAGCGAC	66	Reverse, HisBh35_ mcs1_pACYC_ plasmid overlap
CCTCTGGCTCTTTAAggcgcgccaggagataccATGTTATCCGAACGAACAAAACTAG	63	Forward, Bh33,34 overlap35_ mcs1_pACYC_ AscI
CAAAAATATGCAAGTTTTGCATttcaagttcccctTTAACCCATGAGCATATCTTTGG	63	Reverse, Bh33,34 overlap36_ mcs1_pACYC_ Bh36 overlap
GCAGACTCATTGATTATTTTCATagtccacctcttTTAACCCATGAGCATATCTTTGG	62	Forward, Bh38_ mcs1_pACYC_ Bh34 overlap
GATATGCTCATGGGTTAAaagaggtggactATGAAAATAATCAATGAGTCTGCAC	62	Reverse, Bh33,34_ mcs1_pACYC_ Bh38 overlap
GCAAGAGATAACTTCCAAAGATATGCTCATGGGTTAAaaggggaacttgaaATGCAAAAC	64	Forward, Bh36,7,8_mcs1_pACYC_ Bh34 overlap
ctgttcgacttaagcattatgcggccgcCTAAAACCAAAGCTTATGAAAAGGGTAG	64	Reverse, Bh36,7,8_mcs1_pACYC_ NotI
caccacagccaggatccgaattcgagctcggcgcgccacATGCCAACATCGAAACTTG	60	Forward, Bh43,44_mcs1_pET1_ AscI
CTAGATTGTTTCACTGACAACATCACCTTCTAACTTTTTCATAAATGTC	60	Reverse, Bh43,44_ mcs1_pET1_ Bh45 overlap
ATTGGGGATTAAGGAGGTTTCTGAACATGCGTGAATTTCCAACAATTG	66	Forward, Bh45,46_ mcs1_pET1_ Bh44 overlap
ctttctgttcgacttaagcattatgcggccgcTTAAATATTAGCTTGAAGCGTTTCTTGC	66	Reverse, Bh45,46_ mcs1_pET1_ NotI
gttaagtataagaaggagatatacatatgATGATCATTAGCATTTCAAATCTCTCAC	65	Forward, Bh47-50_mcs2_pET1_ NdeI
gcagcagcggtttctttaccagactcgagTTATTTGCTTTGATATCGTAACGAGATTTC	65	Reverse, Bh47-50_ mcs2_pET1_ XhoI
AATTCCTGCAGTAATACGACTCACTATAGGGGCTATAGCTCAGCTGGGAGAGCGC		$tRNA^{Ala}$_F
mUmGGTGGAGCTATGCGGGATCGAACCGCAGACCTCCTGCGTGCAAAGCAGGCGCTCTCCC		$tRNA^{Ala}_{TGC}$_R
AATTCCTGCAGTAATACGACTCACTATAAGGGGCGTAGTTCAATTGGTAGAGCACCG		$tRNA^{Trp}$_F
mUmGGCAGGGGCGGAGAGACTCGAACTCCCAACACCCGGTTTTGGAGACCGGTGCTCTA		$tRNA^{Trp}$_R

m stands for 2'-O-methyl sugar

Small scale expressions of portions of the *hdn* gene cluster were carried out as described in Chapter 4. Larger expressions were carried out as described in Chapter 2 but with a post-induction temperature of 18 °C rather than 37 °C.

5.4.4 Expression and purification of sLanB proteins

Expressions and purifications were carried out as described in Chapter 2 for LanM proteins (12), with the following modifications. HdnB Lysis Buffer (20 mM Tris, pH 7.6, 500 mM NaCl, 10% glycerol), HdnB Elution Buffer (20 mM Tris, pH 7.6, 500 mM NaCl, 500 mM imidizole, 10% glycerol), HdnB SEC Buffer (20 mM HEPES, pH 7.5, 600 mM KCl, 10% glycerol). Rather than PD-10 size exclusion chromatography, all proteins were loaded onto a Superdex-200 gel filtration column equilibrated with SEC Buffer in order to minimize the phosphate contamination. Proteins were concentrated using an Amicon centrifugal filter with a 30 kDa nominal molecular weight limit (NMWL, EMD Millipore).

5.4.5 In vitro transcription and purification of tRNA

E. coli tRNA and aminoacyl-tRNA synthetase (aaRS) were used since full conversion of HdnA by HdnB1 and HdnA+Ala by HdnB7 was observed using these co-factors. Preparation of tRNA was performed as previously and briefly described below (1). In order to ensure good yield, all buffers were freshly prepared (using RNase free or autoclaved water) for the generation of template DNA or the in vitro transcription reactions. First, double-stranded template DNA was generated by incubating final concentrations of 1× NEB Buffer 2, 4 μM of each primer (see above primer table), 33 μM of dNTPs, and 1 U/μg of DNA of NEB DNA pol I (large Klenow fragment) in a final volume of 50 μL. Several reactions (rather than larger volume reactions) were setup to obtain sufficient quantities of double-stranded DNA. Reactions were incubated at 25 °C for 15-30 min before being quenched with EDTA (10 mM final concentration) and heating at 75 °C for 20 min. The resulting DNA was precipitated using ethanol (2 volumes of pre-chilled ethanol was added and the sample was incubated at −20 °C for 30-180 min), dried using nitrogen, redissolved in water, and used in transcription reactions. Transcription reactions were setup according to a previously described method (13). The resulting tRNA was purified by acidic phenol extraction (for tRNA precipitation, 2.5 volumes of pre-chilled ethanol was added

and the sample was incubated at −20 °C for 1 h) and denaturing, acidic polyacrylamide gel as previously described for aa-tRNA (10). Concentration of tRNA was achieved by using a 10 kDa NMWL Amicon centrifugal filter according to the manufacturer's instructions.

5.4.6 In vitro sLanB reactions and preparation of ^{31}P NMR samples

In vitro sLanB reactions consisted of final concentrations of 200 mM HEPES, pH 7.6, 1 mM ATP, 10 mM $MgCl_2$, 100 mM KCl, 8 µM amino acid, 6 µM tRNA, 1 µM aaRS, 20 µM peptide, and 8 µM HdnB enzyme. The inclusion of amino acid, tRNA, and aa-RS, allowed for in situ aminoacylation due concerns regarding the generation and purificaiton of sufficient amounts of aa-tRNA. In order to produce sufficient quantities of HdnA+Ala-^{18}O and $^{18}OPO_3^{3-}$ (for detection by NMR), HdnB reactions were run on a 1-2 mL scale. Isotopically labelled HdnA+Ala-^{18}O was prepared using the same in vitro reaction conditions, except Ala was substituted with Ala-$^{18}O_2$. The progress of the in vitro reactions was analyzed by subjecting the sample to Zip-Tip and MALDI-TOF-MS (details in Chapter 2, Section 2.4). Full conversion of HdnA to HdnA+Ala was observed after 15 min, thus, reactions were allowed to incubate for 30-60 min at 25 °C. After confirming full conversion to HdnA+Ala-^{18}O, the isotopically labelled peptide was isolated from the reaction mixture using analytical scale RP-HPLC (described in Chapter 2, Section 2.4).

The HdnB7 in vitro reactions were setup and monitored in the same way described for the HdnB1 reactions. However, after incubation, the sample was passed through a 0.5 mL Amicon centrifugal filter with a 10 kDa NMWL to remove the proteinaceous components, and the reservoir was rinsed with two additions of 100 µL of water. The total flow-through was flash frozen and dried to completion *in vacuo*. The solid was re-dissolved in 150-180 µL of D_2O and insoluble components were removed by centrifugation. The resulting supernatant was transferred into a 3-5 mm NMR tube. A 600 MHz Agilent NMR maintained by the Carl R. Woese Institute for Genomic Biology at UIUC was used to acquire spectra. Data was acquired using VnmrJ and analyzed using MestReNova. The identity of phosphate-containing components (e.g. ATP, ADP) was confirmed by spiking using stock solutions to a final concentration of 0.2-1 mM.

17X-PTDH was used to generate $^{18}OPO_3^{3-}$ and PO_4^{3-} by incorporating H_2O^{18} into reactions based on previous conditions (14) and described as follows. The following components were combined to a final volume of 500 µL and incubated at 25 °C for 2.5 h (100 mM MOPS,

pH 7.25, 3.77 mM NAD^+, 1 mM HPO_3^{3-}, 2 µM 17X-PTDH, 27 µL of H_2O^{18} , and 46 µL of H_2O). After preparation and sample analysis as described in the preceding paragraph, a doublet was observed at 2.5 p.p.m (~2.78 mM total phosphate). Dilution of the sample (~10-fold) resulted in loss of resolution and signal. Phosphate contamination was observed in tRNA that had been not been subjected to PAGE purification.

5.5 References

1. Ortega, M.A., Hao, Y., Zhang, Q., Walker, M.C., van der Donk, W.A., and Nair, S.K. (2015). Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature*. 517, 509-512.

2. Ortega, M.A., Hao, Y., Walker, M.C., Donadio, S., Sosio, M., Nair, S.K., and van der Donk, W.A. (2016). Structure and tRNA specificity of MibB, a lantibiotic dehydratase from *Actinobacteria* involved in NAI-107 biosynthesis. *Cell Chem. Biol.* 23, 370-380.

3. Hudson, G.A., Zhang, Z., Tietz, J.I., Mitchell, D.A., and van der Donk, W.A. (2015). In vitro biosynthesis of the core scaffold of the thiopeptide thiomuracin. *J. Am. Chem. Soc.* 137, 16012-16015.

4. Koehnke, J., Bent, A.F., Zollman, D., Smith, K., Houssen, W.E., Zhu, X., Mann, G., Lebl, T., Scharff, R., Shirran, S., Botting, C.H., Jaspars, M., Schwarz-Linek, U., and Naismith, J.H. (2013). The cyanobactin heterocyclase enzyme: a processive adenylase that operates with a defined order of reaction. *Angew. Chem. Int. Ed.* 52, 13991-13996.

5. Dunbar, K.L., and Mitchell, D.A. (2013). Insights into the mechanism of peptide cyclodehydrations achieved through the chemoenzymatic generation of amide derivatives. *J. Am. Chem. Soc.* 135, 8692-8701.

6. Cohn, M. (1982). ^{18}O and ^{17}O effects on ^{31}P NMR as probes of enzymatic reactions of phosphate compounds. *Annu. Rev. Biophys. Bioeng.* 11, 23-42.

7. Ulrich, E.C., and van der Donk, W.A. (2016). Cameo appearances of aminoacyl-tRNA in natural product biosynthesis. *Curr. Opin. Chem. Biol.* 35, 29-36.

8. Moutiez, M., Belin, P., and Gondry, M. (2017). Aminoacyl-tRNA-utilizing enzymes in natural product biosynthesis. *Chem. Rev.* DOI: 10.1021/acs.chemrev.6b00523.

9. Jordan, P.A., and Moore, B.S. (2016). Biosynthetic pathway connects cryptic ribosomally synthesized posttranslationally modified peptide genes with pyrroloquinoline alkaloids. *Cell. Chem. Biol.* 23, 1504-1514.

10. Walker, S.E., and Fredrick, K. (2008). Preparation and evaluation of acylated tRNAs. *Methods.* 44, 81-86.

11. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods.* 6, 343-341.

12. Li, B., Cooper, L.E., and van der Donk, W.A. (2009). Chapter 21. In vitro studies of lantibiotic biosynthesis. *Methods Enzymol.* 458, 533-558.

13. Rio, D.C., Ares, M.J., Hannon, G.J., and Nilsen, T.W. (2011). RNA: a laboratory manual. *Cold Spring Harbor Laboratory Press.*, 216–219.

14. Woodyer, R., van der Donk, W.A., and Zhao, H. (2003). Relaxing the nicotinamide cofactor specificity of phosphite dehydrogenase by rational design. *Biochemistry.* 42, 11604-11614.